



Center for applied Statistics and Economics  
Humboldt-Universität zu Berlin

# Analyse der Berufsvorbereitung während der Schulzeit an Gymnasien und Integrierten Sekundarschulen in Berlin

Analysis of preparation for the job life during school time at two different  
types of high schools in Berlin



## Bachelorarbeit

zur Erlangung des akademischen Grades *Bachelor of Science*  
in Betriebswirtschaftslehre

vorgelegt von

**Nico Keskin**

Matrikelnummer: 546125

Abgabedatum: 14. Juli 2014

Prüfer: Prof. Dr. W. Härdle  
Betreuer: Dr. S. Klinke

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>III</b>
<b>Abkürzungen</b>	<b>IV</b>
<b>1. Einleitung</b>	<b>1</b>
<b>2. Hintergrundinformationen zum Berliner Schulsystem und zur Umfrage</b>	<b>4</b>
2.1. Das Berliner Schulsystem .....	4
2.2. Ziele der Strukturreform .....	6
2.3. Schülerauswahl und Aufbau des Fragebogens .....	9
<b>3. Datenanalyse</b>	<b>11</b>
3.1. Deskriptive Datenanalyse .....	12
3.2. Multivariate Datenanalyse .....	13
3.2.1. Theorie .....	13
3.2.1.1 Faktorenanalyse .....	14
3.2.1.2 Clusteranalyse .....	22
3.2.2. Anwendung auf den vorliegenden Datensatz .....	34
3.2.2.1 Faktorenanalyse .....	35
3.2.2.2 Clusteranalyse .....	42
3.3. Weitere Analysen .....	50
3.4. Repräsentativität der Umfrage .....	53
<b>4. Zusammenfassung</b>	<b>54</b>
<b>Anhang</b>	<b>57</b>
A.1. Literatur- und Quellenverzeichnis .....	57
A.2. Verzeichnis der Variablenabkürzungen .....	59
A.3. SPSS-Output .....	60
A.4. Muster-Fragebogen .....	73

# Abbildungsverzeichnis

Abbildung 2.1.:	Bildungsfahrplan Berlin: <i>Senatsverwaltung für Bildung, Wissenschaft und Forschung - Bildungsfahrplan S. 6</i> (Seite 4)
Abbildung 3.1.:	Beispiel einer Faktorenmatrix mit Eigenwerten und Kommunalitäten (Seite 19)
Abbildung 3.2.:	Screen-Test Beispiel (Seite 20)
Abbildung 3.3.:	Muster einer Rohdatenmatrix (Seite 23)
Abbildung 3.4.:	Muster einer Distanz- oder Ähnlichkeitsmatrix (Seite 24)
Abbildung 3.5.:	Paarweiser Vergleich von binären Variablen (Seite 25)
Abbildung 3.6.:	Berechnung der Distanzen bei hierarchischen, agglomerativen Verfahren (Seite 30)
Abbildung 3.7.:	KMO-Kriterium (Seite 35)
Abbildung 3.8.:	Kommunalitäten-Matrix der angewandten Extraktionsverfahren (Seite 36)
Abbildung 3.9.:	Eigenwerte der Extraktionsverfahren mit und ohne Rotation (Seite 37)
Abbildung 3.10.:	Screen-Test (Seite 38)
Abbildung 3.11.:	Faktorladungsmatrix (Seite 39)
Abbildung 3.12.:	Komponentendiagramm im rotierten Raum (Seite 41)
Abbildung 3.13.:	Dendrogramm Single-Linkage (Seite 43)
Abbildung 3.14.:	Zusammenfassung Ward-Verfahren (Seite 43)
Abbildung 3.15.:	Dendrogramm Ward-Verfahren (Seite 44)
Abbildung 3.16.:	Test von Mojena (Seite 45)
Abbildung 3.17.:	Dendrogramm Ward-Verfahren mit markierten Clustern (Seite 45)
Abbildung 3.18.:	Clusteranalyse: Signifikant-unterschiedliche Variablen (Seite 46)
Abbildung 3.19.:	Unterschiede nach Schultypen (Seite 51)

# Abkürzungen

BB	Berufsbildungsreife (früher: Hauptschulabschluss)
eBB	Erweiterte Berufsbildungsreife (früher: erweiterter Hauptschulabschluss)
EFA	Explorative Faktorenanalyse
ISS	Integrierte Sekundarschule
KMO	Kaiser-Meyer-Olkin-Kriterium
MSA	Mittlerer Schulabschluss (früher: Realschulabschluss)
MSA	Measure of Sampling Adequacy
OSZ	Oberstufenzentrum
PAF	Hauptachsenanalyse
PCA	Hauptkomponentenanalyse
SPSS	Statistical Package for the Social Sciences (für die Analyse der erhobenen Daten benutzte Statistiksoftware)
Teiln.	Teilnehmer

# 1. Einleitung

Die Bedeutung von Bildung wird allein daraus ersichtlich, dass sie in vielen Ländern der Erde als Grundrecht eines jeden Menschen angesehen wird. Schon in dem “Internationalen Pakt über wirtschaftliche, soziale und kulturelle Rechte” aus dem Jahre 1966 schreiben die Vereinten Nationen als eine der führenden Menschen- und Völkerrechtsorganisationen mit weltweit 193 Mitgliedstaaten: „Die Vertragsstaaten erkennen das Recht eines jeden auf Bildung an. Sie stimmen darin überein, dass Bildung auf die volle Entfaltung der menschlichen Persönlichkeit und des Bewusstseins ihrer Würde gerichtet sein und die Achtung vor den Menschenrechten und Grundfreiheiten stärken muss. Sie stimmen ferner überein, dass Bildung es jedermann ermöglichen muss, eine nützliche Rolle in einer freien Gesellschaft zu spielen, dass sie Verständnis, Toleranz und Freundschaft unter allen Völkern und allen rassischen, ethnischen und religiösen Gruppen fördern sowie die Tätigkeit der Vereinten Nationen zur Erhaltung des Friedens unterstützen muss.”<sup>1</sup> Des Weiteren wird das Recht auf Bildung in dem “Übereinkommen über die Rechte des Kindes” (Vereinte Nationen, 1989), sowie in der “Charta der Grundrechte der Europäischen Union” in Artikel 149 geregelt.

In der Bundesrepublik Deutschland ist das Recht auf Bildung indirekt im Grundgesetz verankert. Implizit lässt es sich aus den Grundrechten der Menschenwürde und der Gleichberechtigung ableiten, denn niemandem darf der Zugang zu Bildung verwehrt werden und niemand darf beim Erlangen der Bildung diskriminiert werden.<sup>2</sup> Den Großteil an Bildung, den ein Mensch in Deutschland erlangt, wird in der Schule vermittelt. Mit dem Wissen um die Bedeutung von Bildung und der Aufgabe der Vermittlung von Werten haben die Schulen einen besonders hohen Stellenwert in unserer Gesellschaft.

Zu den wichtigsten Aufgaben<sup>3</sup> der Schule zählen die Betreuung der Schüler (kustodiale Funktion), die Integrationsfunktion, die Qualifikationsfunktion und die Selektionsfunktion. Integrationsfunktion bedeutet, dass sich die Schüler möglichst reibungs-

---

<sup>1</sup> Vereinte Nationen - Internationaler Pakt über wirtschaftliche, soziale und kulturelle Rechte (1966): Artikel 13

<sup>2</sup> Grundgesetz für die Bundesrepublik Deutschland: Artikel 1, Artikel 3 (1949)

<sup>3</sup> Hintz, D.; Pöppel, K.; Rekus, J.: S. 118 - 123

los in das Gesellschaftsleben eingliedern. Dafür sollen den Schülern Normen und Werte vermittelt werden, sodass sie zur Stabilität der Gesellschaft beitragen. Außerdem soll eine Sensibilisierung für Tugenden im Umgang mit Sachen (Genauigkeit, Sparsamkeit), im sozialen Umgang (Gehorsam, Verlässlichkeit, Pünktlichkeit) und in den Bereichen Gerechtigkeit, Ordnung, Zeit und Arbeit<sup>4</sup> erreicht werden.

Die Qualifikationsfunktion bezieht sich auf das Vermitteln von Fähigkeiten und Kenntnissen, um jeden Einzelnen mit bestmöglichem Wissen auszustatten und somit zu einem eigenständigen Leben zu befähigen. Durch technologische Innovation und kulturelle Veränderungen muss es dabei ständig zu Anpassungen des Unterrichts kommen. Nur so kann eine optimale Vorbereitung auf die sich verändernde Arbeitswelt garantiert werden. So zählen heutzutage das Beherrschen mindestens einer Fremdsprache und der Umgang mit einem Computer zu den Standards für das Erlangen eines Schulabschlusses. Derzeit wird bei Schulabgängern immer mehr Wert gelegt auf das Vorhandensein von Schlüsselqualifikationen (Soft-Skills), wie beispielsweise Lern- und Kommunikationsbereitschaft und Entscheidungskompetenz. Eben solche fachlichen und persönlichen Kompetenzen sind aus Sicht der Industrie- und Handelskammer die wichtigsten Eigenschaften, die die Wirtschaft von Schulabgängern erwartet<sup>5</sup>.

Die Selektionsfunktion bezeichnet die Umverteilung der Schüler auf verschiedene Schullaufbahnen und Lebenschancen. Noten und verschiedene Abschlüsse dienen dabei als Steuerungsmechanismen. Dieser Vorgang entscheidet im Endeffekt über den Erfolg im Berufsleben, dem sozialen Ansehen und dem materiellen Wohlstand. Trotz aller Bemühungen, verschiedene Einflüsse auf diesen Allokationsprozess zu eliminieren, haben soziale und regionale Herkunft sowie das Geschlecht immer noch einen Einfluss auf die Verteilung der Lebenschance, und nicht nur die persönliche Begabung und vielleicht noch ein nötiges Quäntchen Glück sind die einzigen Einflussfaktoren.

Wie bereits erwähnt, sind die Schulen durch die Vermittlung der nötigen fachlichen und persönlichen Kompetenzen maßgeblich für die Vorbereitung der Schüler auf das Berufsleben mitverantwortlich. Dabei bedeutet Vorbereitung auch, den Schülern bei

---

<sup>4</sup> Fend, H. (1980)

<sup>5</sup> Industrie- und Handelskammer: <http://www.frankfurt-main.ihk.de/berufsbildung/marktplatz/schulabgaenger/> (19.05.2014)

der Interessensfindung behilflich zu sein und Unterstützung bei der Auswahl von Berufswünschen zu leisten (z. B. Anleitungen für Bewerbungsschreiben oder Vorbereitungen für Bewerbungsgespräche).

Doch inwiefern lassen sich anhand der erhobenen Daten die Berufsvorbereitung beeinflussende Faktoren feststellen bzw. homogene Gruppen von Schülern bilden? Diese Fragen hoffe ich am Ende meiner Arbeit, in der ich insbesondere die Selektions- und Qualifikationsfunktion der Schulen betrachte, beantworten zu können, sowie Unterschiede in der Berufsvorbereitung an den weiterführenden Schultypen aufdecken zu können. An dieser Stelle möchte ich darauf hinweisen, dass ich mich in dieser Arbeit ausschließlich mit den zwei allgemeinbildenden, weiterführenden Berliner Schularten - dem Gymnasium und der Sekundarschule - beschäftigte. Die beruflichen Bildungswege (OSZ, Berufsschule, usw.) finden an dieser Stelle keine Beachtung.

Zur Datenerhebung habe ich eine Umfrage an zwei Schulen mit Hilfe eines selbst erstellten Fragebogens durchgeführt. Die Teilnehmer mussten Fragen zur Berufsvorbereitung in der Schule und privat, sowie zu ihren Neigungen, Vorlieben und Charakterzügen beantworten<sup>6</sup>. Im Rahmen der Datenauswertung erhoffe ich mir, u. U. gleichermaßen vorliegende Muster aber auch Unterschiede zwischen den Schulformen herausarbeiten zu können. Des Weiteren wäre interessant, ob eventuell vorliegende Cluster nachgewiesen werden können. Zusätzlich werden die erhobenen Daten einer Faktorenanalyse unterzogen, um ggf. vorhandene Muster in der Berufsvorbereitung schulübergreifend zu ermitteln.

Um die erhaltenen Ergebnisse korrekt einschätzen und interpretieren zu können, beginne ich unter *Punkt 2* mit einer theoretischen Einführung in das Berliner Schulsystem, einer Erläuterung der Strukturreform, die in den letzten Jahren vollzogen wurde, sowie mit einer Erklärung zum Aufbau der Umfrage und des Fragebogens. Im darauf folgenden Abschnitt, *der Datenanalyse*, wird der Datensatz anhand einiger deskriptiven Mittel beschrieben. Anschließend werden die Konzepte der Faktoren- und Clusteranalyse vorgestellt, die dann auf den Datensatz angewendet werden. Zum Abschluss werde ich unter *Punkt 4* alle wichtigen Schlussfolgerungen nochmals gebündelt darstellen und in einen Zusammenhang bringen. Im Anhang werden die Quellen,

---

<sup>6</sup> Weitere Erläuterungen zum Fragebogen unter 2.3.

ein Variablenverzeichnis, ein Muster-Fragebogen und alle zum Nachvollziehen der Ergebnisse notwendigen SPSS-Outputs aufgelistet.

Im weiteren Verlauf des Textes wird der Ausdruck "Schülerinnen und Schüler" zur Verbesserung des Leseflusses durch "Schüler" abgekürzt.

## 2. Hintergrundinformationen zum Berliner Schulsystem und zur Umfrage

### 2.1. Das Berliner Schulsystem

Die Berliner Schullandschaft wurde in den letzten Jahren, vor allem durch die Berliner Schulstrukturereform, die zum Schuljahr 2010/11 in Kraft trat, grundlegend verändert. Die Vorbereitungen starteten mit dem Schuljahr 2009/2010 und die Implementierung der erlassenen Reformen soll nach 5 Jahren abgeschlossen sein. Die größte Veränderung besteht in dem Komprimieren von ehemals fünf verschiedenen weiterführenden Schularten in zwei alternative Auswahlmöglichkeiten. Die ehemaligen Hauptschulen, Realschulen, verbundenen Haupt- und Realschulen und Gesamtschulen wurden zu integrierten Sekundarschulen (ISS) zusammengefasst. Das Gymnasium in seiner bisherigen Form wurde beibehalten. Damit einhergehend sollen alle integrierten Sekundarschulen - ähnlich wie schon die Grundschulen - zu Ganztagschulen umstrukturiert werden. Für Gymnasien wurde vorgegeben, dass es pro Bezirk zunächst mindestens eine Schule mit Ganztagsangebot geben soll.

Ein wichtiges Kriterium dieser Reform ist, dass an beiden Schultypen die gleichen Bildungsabschlüsse erlangt werden können, d.h. auch an einer ISS kann am Ende der Schulausbildung das Abitur erworben werden.

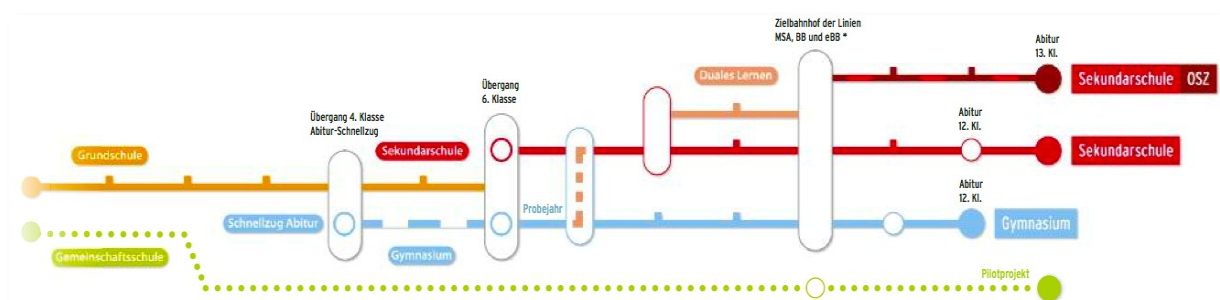


Abbildung 2.1.: Bildungsfahrplan Berlin



Wie der Abbildung 2.1. zu entnehmen ist, hat jeder Schüler nach der Grundschule die Möglichkeit, entweder eine integrierte Sekundarschule oder ein Gymnasium zu besuchen. Diese Entscheidung wird seit der Schulreform alleine von den Eltern getroffen und ist nicht mehr von den Empfehlungen der Schulen abhängig. Die Abbildung zeigt außerdem, dass es weiterhin die Möglichkeit gibt, an bestimmten Gymnasien ab der 5. Klasse über die Schnellläufer-Schiene das Abitur in wahlweise 11 oder 12 Jahren zu erlangen. Wer das Gymnasium ab der 7. Klasse besucht, muss sein Abitur nach der 12. Klasse ablegen. Das an einer ISS erworbene Abitur ist mit dem Abitur am Gymnasium gleichwertig. Dabei bieten integrierte Sekundarschulen entweder selbst eine gymnasiale Oberstufe an oder kooperieren mit einer solchen ISS oder einem Oberstufenzentrum (OSZ). Wer das Abitur über eine ISS ablegen möchte, kann dies wahlweise in Abhängigkeit der Leistungen nach 12 oder 13 Jahren tun; in Kooperation mit einem OSZ sind 13 Jahre jedoch Pflicht.

Weitere, weniger qualifizierte Bildungsabschlüsse sind der MSA, die Berufsbildungsreife und die erweiterte Berufsbildungsreife. Auch diese beiden Abschlüsse können an beiden Schultypen erworben werden. Für die Erlangung des MSAs ist ein zentraler Prüfungskomplex, bestehend aus den Fächern Deutsch, Mathematik und der ersten Fremdsprache, sowie einer vierten Prüfungskomponente, in der die Schüler ein Thema ihrer Wahl präsentieren müssen, zu bewältigen. Welcher der drei Abschlüsse erreicht wird, hängt zum Einen vom Bestehen der zentralen Prüfungen ab und zum Anderen von den sonstigen Noten im regulären Schulbetrieb. Um von einer ISS in die gymnasiale Oberstufe wechseln zu können, muss einerseits der MSA bestanden sein, ein bestimmter Notendurchschnitt erreicht werden und eine Anzahl von Leistungskursen nachgewiesen werden.

Am Ende der Oberstufenzeit müssen an beiden Schultypen die Abiturprüfungen abgelegt werden, bestehend aus jeweils einer Prüfung in den gewählten Leistungskursen, einer Prüfung in einem Grundkurs, einer mündlichen Prüfung und einer Präsentation über ein frei zu wählendes Thema.

Durch das Verkürzen der Schulzeit für Abiturienten um ein Jahr kommt es zu durchschnittlich längeren Schultagen für Gymnasiasten, da die Menge des zu vermittelnden Stoffs unverändert bleibt. Im Mittel haben Schüler eines Gymnasiums 2 Stunden mehr Unterricht pro Woche als Schüler einer ISS. Aufgrund der unterschiedlichen

Konzepte der beiden Schultypen haben Schulklassen einer ISS mit einer durchschnittlichen Schülerzahl von 25 im Mittel vier Schüler weniger als Schulklassen eines Gymnasiums. Diese unterschiedlichen Konzepte, sowie das speziell für die ISS entwickelte Konzept des Dualen Lernens werden ausführlich in Abschnitt 2.2. behandelt.

Eine gewisse Sonderstellung, ebenfalls in Abbildung 2.1. dargestellt, hat die sogenannte Gemeinschaftsschule. Dieses Schulmodell läuft seit dem Schuljahr 2008/09 als Pilotprojekt. In der Gemeinschaftsschule bleiben die Klassen von der Einschulung bis zum Schulabschluss nach der 10., 12. oder 13. Klasse fest bestehen, das heißt, auch hier sind alle Abschlüsse möglich.

Des Weiteren bietet das Berliner Bildungssystem noch zusätzlich die Sonderpädagogische Förderung an, die ebenfalls mit den von den allgemeinen Bildungswegen bekannten Schulabschlüssen beendet werden kann.

## 2.2. Ziele der Strukturreform

Im Berliner Bildungswesen war vielen Bildungsexperten zufolge<sup>7</sup> das dreiteilige Bildungssystem mit Hauptschule, Realschule und Gymnasium nicht mehr zeitgemäß. Dies lag vor allem daran, dass Hauptschulen oft als Restschulen angesehen wurden und es als sozial nachteilig galt, eine solche Schule zu besuchen. In den vergangenen Jahren haben sich deshalb nur noch ca. 7% der Schüler für den Besuch einer Hauptschule entschieden.

Die neue Integrierte Sekundarschule ermöglicht Schülern, die sowohl die (erweiterte) Berufsbildungsreife, den MSA oder auch das Abitur anstreben, gemeinsam zu lernen und am Unterricht teilzunehmen. Dieses Modell des gemeinsamen Lernens orientiert sich an den Schulstrukturen der skandinavischen Länder, die dieses Schulsystem schon lange praktizieren und regelmäßig vordere Plätze im PISA-Ranking einnehmen. Durch den Umbau der ISS zu kostenlosen Ganztagschulen erhofft man sich folgende Effekte aufgrund des größeren zur Verfügung stehenden Betreuungszeitfensters:

---

<sup>7</sup> Senatsverwaltung für Bildung, Wissenschaft und Forschung - Bildungsfahrplan S. 2

- Förderung der Zusammenarbeit der Schulen untereinander
- Verbesserung der Zusammenarbeit mit externen Unternehmen, Verbänden und Sozialarbeitern
- Verbesserung der individuellen Förderung und Betreuung von Jugendlichen mit dem Ziel der Verbesserung der Chancengleichheit

Besonders von Interesse bei der Schulreform war die Verbesserung der Integration von Schülern mit Migrationshintergrund und von Schülern aus sozial schwachen Familien, denn wie internationale Studien<sup>8</sup> zeigen, ist der Erfolg in der Schule häufig an die soziale Herkunft der Schüler geknüpft. Durch das Einführen des Systems der Ganztagschule an den ISS und die bessere Betreuung bis in den Nachmittag hinein, liegt die Hoffnung nahe, den Lernerfolg speziell dieser Schüler effektiv zu verbessern. Zudem werden neben dem Unterrichtsbetrieb zusätzliche Angebote zur Freizeitgestaltung, Integration oder Berufsvorbereitung bereitgestellt. Beispielsweise gibt es bestehende Kooperationen mit Sportvereinen, Musik- und Volkshochschulen, Werkstätten, usw..

Etwaige Unterschiede im Bildungsniveau der Schüler von Gymnasien und den ISS werden durch verschiedene Mechanismen ausgeglichen. An einer ISS kann zum Beispiel gewählt werden, ob man sein Abitur nach 12 oder 13 Jahren ablegen möchte, abhängig von der persönlichen Leistungsfähigkeit, während an Gymnasien 12 Jahre Pflicht sind oder Schüler die Schnellläufer-Schiene in 11 Jahren durchlaufen. Zwangsläufig unterscheiden sich dadurch beide Schulmodelle in der Anzahl der durchschnittlichen Schulstunden pro Woche bei gleichem zu vermittelnden Stoff. Zusätzlich ist die Anzahl an Schülern pro Klasse bzw. Kurs geringer. Dadurch kann den Sekundarschülern eine bessere individuelle Förderung und ein nachhaltigeres Lernen angeboten werden. Ferner erhalten in den ISS Klassen mit einem hohem Anteil an Schülern mit nichtdeutscher Herkunft zusätzliche Lehrkräfte, um eine angemessene Betreuung zu gewährleisten.

Auch an den Lehrmethoden selbst wurden einige Veränderungen vorgenommen. Die Tendenz geht weg von dem klassischen Frontalunterricht hin zu individuelleren Lehrmethoden. Das modernere Lernen soll sich an die unterschiedlichen Lerngeschwindigkeiten, Lernstile und Interessen der Schüler anpassen und vor allem soll

---

<sup>8</sup> Senatsverwaltung für Bildung, Wissenschaft und Forschung - Bildungsfahrplan S. 2

eigenständiges Lernen vermittelt werden. Kennzeichnend für die ISS ist der differenzierte Unterricht in den einzelnen Fächern. Die Kurse können in Abhängigkeit der Leistungsnachweise gewechselt werden. Zusätzlich bieten die ISS Klassen mit besonderen Schwerpunkten an (z.B. bilinguale, naturwissenschaftliche, künstlerische oder musische Klassen).

Bei der Schulreform wurde besonders auf die Verbesserung der Berufsvorbereitung der Schüler sowohl an Gymnasien als auch an ISS geachtet. Dazu wurde das sogenannte Duale Lernen insbesondere an den ISS in den Unterrichtsbetrieb integriert. Da die Entscheidungsgewalt über spezielle Maßnahmen bei den einzelnen Schulen liegt, ist auch für Gymnasien das Konzept des Dualen Lernens möglich. Das Duale Lernen kann neben dem üblichen Schulbetrieb aus Pflichtveranstaltungen wie beispielsweise Betriebspraktika oder Karrierewochen und weiteren freiwilligen Veranstaltungsteilnahmen bestehen. Andere Anwendungsgebiete sind die Kooperation mit Berufsberatungen, Ausbildungszentren, Werkstätten, Hochschulen und Unternehmen, um den Schülern die bestmögliche Orientierung beim Einstieg in die Berufswelt zu ermöglichen. Außerdem kann so erreicht werden, dass Schüler bereits in jungen Jahren bei der Interessensfindung begleitet werden. Grundsätzlich ist das Duale Lernen für die Jahrgangsstufen 7 bis 10 vorgesehen. Für die höheren Jahrgänge lassen sich Interessen durch die Zusammenarbeit mit Hochschulen für Studieninteressierte oder Aktionen wie Jugend forscht wecken. Zusätzlich erhofft man sich, dass abschlussgefährdete Schüler infolge mangelnder Lern- und Leistungsbereitschaft durch regelmäßiges Teilhaben an der Berufswelt Perspektiven aufgezeigt bekommen, so dass es zu neuer Motivation kommt, die Schule erfolgreich zu beenden.

Alles in allem soll die Schulreform die individuelle Förderung der Schüler verbessern, sowie das Zusammenleben und -lernen unterschiedlicher Schüler optimieren. Dabei können die Unterschiede sowohl in der sozialen oder regionalen Herkunft als auch in dem Leistungspotential, den Interessen oder den persönlichen Zielen des Einzelnen liegen. Des Weiteren wurde durch die erhöhten Berufsvorbereitungsmaßnahmen den Sekundarschülern und Gymnasiasten die Möglichkeit gegeben, sich schon während der Schulzeit ein Bild über das Berufsleben zu machen und mögliche Interessenschwerpunkte zu entdecken. Wie bereits eingangs erwähnt, wird in der vorliegenden Arbeit untersucht, ob und wie sich die Berufsvorbereitung an den beiden Schultypen

unterscheidet und inwiefern sich Muster bei Berufsvorbereitung oder spezielle schultypbezogene Gruppen erkennen lassen.

## 2.3. Schülerauswahl und Aufbau des Fragebogens

Um an Informationen für die zu bearbeitende Fragestellung zu kommen, habe ich eine Umfrage mit einem selbst entwickelten Fragebogen durchgeführt. Aufgrund der Themenstellung wurden Schüler beider Schultypen befragt.

Bei der Festlegung, welche Jahrgangsstufen befragt werden sollten, konnten der 12. und 13. Jahrgang ausgeschlossen werden. An Gymnasien gibt es keinen 13. Jahrgang und die Schüler der 12. Klassen, egal ob in der Abitur-Prüfungsphase oder nicht, befinden sich unabhängig vom Schultyp nicht mehr im klassenorientierten Unterricht. Durch die Kurssysteme ist es quasi unmöglich, eine angemessene Anzahl von Schülern zu erreichen. In entsprechender Weise gilt das Gleiche für Schüler der 11. Klasse, die ebenfalls nach dem Kurssystem unterrichtet werden. Der Aufwand wäre unverhältnismäßig groß und wegen der anonymen Befragung hätten Mehrfachbefragungen gleicher Schüler nicht vermieden werden können.

Aufgrund dieser vorangestellten Überlegungen bot sich die Jahrgangsstufe 10 beider Schultypen für die Befragung an. Weitere Gründe sprechen für die Auswahl dieses Jahrgangs:

- Der Unterricht wird im Klassenverband abgehalten, fast alle Schüler können dadurch relativ einfach erreicht werden.
- Auch jene Schüler können einbezogen werden, die nach der 10. Klasse ihre Schullaufbahn beenden und somit in höheren Jahrgängen nicht mehr auftauchen würden.
- 10. Klässler an den ISS häufiger aber auch einige an den Gymnasien beenden ihre Schulausbildung und sollten sich deshalb zwangsläufig auf das Berufsleben danach vorbereitet haben. Sie sollten deshalb bezüglich der Fragen sensibilisiert sein.

Die Befragung von Schülern niedrigerer Klassenstufen wurde nicht in Erwägung gezogen, da sich diese Schüler u. U. noch nicht intensiv mit dem Thema der Berufsvorbereitung beschäftigt haben. Erschwerend hinzugekommen wäre, dass für alle Schüler unter dem vollendeten 14. Lebensjahr eine Genehmigung der Eltern für die Befragung hätte vorliegen müssen. Der Ablauf der Befragung wäre kompliziert und eine genügend hohe Rücklaufquote wäre fraglich geworden.

Im Rahmen der praktischen Umsetzung der Befragung wurden die Fragebögen den Klassenlehrern übergeben. Sie übernahmen die Verteilung der Bögen, überwachten das Ausfüllen durch die Schüler und sammelten anschließend die Bögen wieder ein. Durch das Bearbeiten der Bögen während des Unterrichts konnte eine maximale Rücklaufquote erzielt werden.

Schülerbefragungen sind nur zulässig, wenn beim Berliner Senat ein entsprechender Antrag gestellt und ein positiver Bescheid erteilt wurde. In meinem Fall wurde mir die Befragung von Schülern eines Gymnasiums und einer integrierten Sekundarschule genehmigt. Infolgedessen habe ich eine Vollbefragung des 10. Jahrgangs an beiden Schulen durchgeführt. Das Gymnasium unterrichtet sechs 10. Klassen, von denen zwei Schnellläufer-Klassen sind; an der ISS gibt es acht 10. Klassen. Insgesamt wurden 230 Schüler befragt (97 am Gymnasium und 133 an der ISS). Die Anzahl der abgegebenen Fragebögen deutet nicht auf besonderes kleine Klassen hin, sondern ist dem Umstand geschuldet, dass Schüler krank oder auf Klassenfahrt waren. Außerdem war die Befragung freiwillig und einigen war der Aufwand zu hoch. Ich weise bereits an dieser Stelle darauf hin, dass bei dieser Arbeit keine Repräsentativität gegeben ist, worauf ich später noch in einem gesonderten Punkt eingehen werde.

Der Fragebogen<sup>9</sup> beinhaltet vier verschiedene Themengebiete. Zunächst werden persönliche Informationen erfragt, wie zum Beispiel Alter, Geschlecht, Art der besuchten Schule, usw.. Der zweite Komplex befasst sich mit Fragen zur Berufsvorbereitung in oder ausgehend von der Schule. Damit sind beispielsweise Exkursionen im Rahmen von Berufsvorbereitungsmaßnahmen oder die Berufsvorbereitung als Teil des Unterrichts gemeint. Der dritte Komplex beschäftigt sich mit der persönlichen Berufsvorbereitung und Interessensfindung der Schüler. Erfragt wird, inwiefern sich der Schüler außerhalb der Schule selbst mit seiner Zukunft nach der Schulzeit auseinan-

---

<sup>9</sup> Muster-Fragebogen siehe Anhang A.3.

dergesetzt hat. Abschließend werden Fragen gestellt, die auf die Interessen und die Charakterzüge des Schülers schließen lassen, um eventuell vorhandene Muster ableiten zu können.

Die Fragen wurden leicht verständlich formuliert und auf komplizierte Begriffe wurde verzichtet, ggf. wurden Begriffe erläutert<sup>10</sup>. Der überwiegende Anteil der Fragen wurde im Multiple-Choice Verfahren mit einer Vierer-Abstufung (*trifft voll und ganz zu, trifft eher zu, trifft eher nicht zu, trifft überhaupt nicht zu*) gestellt. Auf eine "Mittelstufe" in Form eines Items "weiß ich nicht", "neutral" oder "kann ich nicht beurteilen" wurde bewusst verzichtet, da solche Optionen gerne als Fluchtmöglichkeit<sup>11</sup> verwendet werden und um sich keine Gedanken machen zu müssen. Somit müssen sich die Schüler für eine Ausprägung bzw. Tendenz entscheiden. An dieser Stelle möchte ich anmerken, dass ich gerne einen differenzierteren Fragenkatalog ausgearbeitet hätte. Fragen nach Religion, Migrationshintergrund, Familienverhältnisse, Einkommen der Eltern, Beschäftigung der Geschwister, usw. unterstehen streng dem Datenschutz und hätten ebenfalls die Einverständniserklärung der Eltern vorausgesetzt. Dies hätte vermutlich wiederum die Rücklaufquote der Fragebögen so reduziert, dass eine zielführende Auswertung nicht möglich gewesen wäre.

### 3. Datenanalyse

Nach Erläuterungen zum Befragungsklientel und Beschreibungen zur Ausgangssituation erfolgt nun die Analyse der erhobenen Daten. Zunächst wird die Stichprobe mittels deskriptiver Methoden beschrieben. Anschließend wird mit Hilfe multivariater Verfahren, die zunächst in einem theoretischen Teil erklärt werden, versucht, unbekannte korrelative Strukturen im Datensatz aufzudecken. Mit weiteren Analyseverfahren wird die Berufsvorbereitung der beiden befragten Schulen verglichen. Zum Schluss dieses Abschnittes werde ich genauer auf die Repräsentativität der durchgeführten Umfrage eingehen.

---

<sup>10</sup> Jacob, Heinz, Décieux (2013): S. 121ff. - 13 Gebote zur Formulierung von Fragen

<sup>11</sup> Jacob, Heinz, Décieux (2013): S. 137

### 3.1. Deskriptive Datenanalyse

Zum Einstieg in die Datenanalyse soll unter diesem Punkt die Stichprobe beschrieben werden:

Wie schon eingangs erwähnt, haben 230 Schüler an der Umfrage teilgenommen. Davon besuchen 97 eine 10. Klasse des befragten Gymnasiums und 133 eine 10. Klasse der ISS. Insgesamt sind 51% der Befragten männlich und 49% weiblich. Dabei ist auffällig, dass nur 30% aller weiblichen Umfrageteilnehmer das Gymnasium besuchen. Bei den männlichen Befragten ist der Anteil mit 55% deutlich höher.

Der größte Anteil der Befragten möchte das Abitur ablegen (69%). Das ist dadurch zu erklären, dass das Abitur als Bildungsabschluss an dem Gymnasium der Normalfall ist und nur wenige Schüler ihren Bildungsweg mit einem anderen Abschluss beenden möchten und zusätzlich viele Schüler der ISS das Ziel haben, das Abitur zu erlangen. Auch hier lässt sich auf Grund der zuvor beschriebenen Ungleichgewichte feststellen, dass die Quote der männlichen Befragten, die das Abitur anstreben, deutlich höher ist als die von weiblichen. Dies ist darauf zurückzuführen, dass eben an dem Gymnasium mehr Jungen als Mädchen befragt wurden - 81% der männlichen Teilnehmer würden gerne das Abitur ablegen; bei den weiblichen sind es 56%. Wegen der höheren Gewichtung der weiblichen Befragten an der ISS möchten 31% der Mädchen zunächst den MSA machen und sich später eventuell weiterqualifizieren. Diesen Plan verfolgen nur 11 % der Jungen an der ISS.

Ebenfalls auffällig ist der hohe Anteil an Schülern der ISS, die ihr Abitur ablegen wollen. Mehr als die Hälfte der Befragten (51%) würden zumindest gerne das Abitur als Bildungsabschluss erlangen. Dabei kann natürlich nicht endgültig geklärt werden, ob alle Schüler, die an der ISS nach der 10. Klasse die gymnasiale Oberstufe besuchen wollen, auch wirklich die dazu benötigten Voraussetzungen erfüllen. Entsprechend hoch bei 48% liegt der Anteil an Sekundarschülern, die vorhaben, sich an einer Universität oder einer Fachhochschule einzuschreiben. Bei den Gymnasiasten liegt der Abiturabschluss wie erwartet sehr hoch bei 93%. An dem Gymnasium beträgt der Anteil der Schüler, die einen Universitätsabschluss anstreben, 89%.

Über die gesamte Stichprobe betrachtet, lässt sich feststellen, dass 65% aller Befragten den Plan haben, eine Uni zu besuchen, 24% wollen eine Ausbildung beginnen,



1% der Befragten plant einen direkten Berufseinstieg und 4% haben vor, einen "sonstigen" Weg einzuschlagen.

Auf einer Schulnotenskala von 1 bis 6 liegt die Durchschnittsbewertung der Schüler bezüglich der persönlichen und schulischen Universitäts- bzw. Berufsvorbereitung bei 3,14. Jeder mit Schulnoten vertraute weiß, dass es sich dabei um eine mittelmäßige Bewertung handelt. Dabei gibt es Unterschiede zwischen Jungen (3,05) und Mädchen (3,21) und auch bei den Schulformen - Gymnasium (3,28) und ISS (3,05). Wie stark diese Unterschiede sind, wird in dem Abschnitt 3.4. genauer untersucht.

Abschließend kann noch festgestellt werden, dass Englisch, Biologie und Deutsch die drei Fächer sind, die die Schüler am liebsten als Leistungskurse wählen würden. Die beliebtesten Branchen für die zukünftige Berufswahl sind die Gesundheitsbranche, das Eventmanagement und die Medien. Die am meisten genannten Interessensgebiete aller Befragten sind Sport, Videospiele und Freunde.

Alle hier genannten Prozentwerte lassen sich im SPSS-Output im Anhang A.3.a nachvollziehen.

## 3.2. Multivariate Datenanalyse

### 3.2.1. Theorie

Der Name lässt erkennen, dass sich die Auswertungen multivariater Analyseverfahren auf viele Variablen gleichzeitig beziehen. Während bei bivariaten Analysen zum Beispiel die Korrelation zwischen Variable 1 und Variable 2 berechnet wird, werden in der multivariaten Analyse die Korrelationen zwischen mehreren Variablen untereinander analysiert. Aufgrund des dadurch bedingten hohen Aufwandes - bei 10 Variablen müssen schon 45 Korrelationen berechnet werden -, werden multivariate Verfahren häufig mit Statistikprogrammen durchgeführt. In dieser Arbeit wurde dafür die Statistiksoftware SPSS von IBM genutzt.

Grundsätzlich kann man die multivariaten Analyseverfahren in struktur-prüfende Verfahren und struktur-entdeckende Verfahren einteilen<sup>12</sup>. Dabei zielen struktur-prüfende Verfahren darauf ab, schon vorab angenommene oder bekannte Zusammenhänge zwischen vorliegenden Variablen zu überprüfen und ggf. zu bestätigen. Im Gegenteil dazu werden struktur-entdeckende Verfahren genutzt, um neue Zusammenhänge zwischen den Variablen des Datensatzes aufzudecken. Beide in dieser Arbeit angewandten Verfahren, die Faktoren- und Clusteranalyse zählen zu den struktur-entdeckenden Verfahren, da für den speziell für diese Arbeit erhobenen Datensatz keine Theorien oder Annahmen vorab aufgestellt wurden.

Beide Verfahren werden im Folgenden anhand des Lehrbuches “Multivariate Analysemethoden” von Backhaus, Erichson, Plinke und Weiber (13. Auflage, 2011) vorgestellt.

### 3.2.1.1. Faktorenanalyse

Bei der Faktorenanalyse unterscheidet man die konfirmative Faktorenanalyse (KFA) und die explorative Faktorenanalyse (EFA). Während bei der konfirmativen Faktorenanalyse „Hypothesen über die Faktorenstruktur eines Datensatzes getestet werden können“<sup>13</sup>, hat die explorative Faktorenanalyse das Ziel, aus „wechselseitigen Beziehungen vieler Variablen ein [neues] Erklärungsmodell“<sup>14</sup> zu entwickeln. Für die vorliegende Arbeit kommt die explorative Faktorenanalyse zur Anwendung, da im Vorfeld der Analyse keine Hypothesen hinzugezogen oder neu aufgestellt wurden. Mit Hilfe der EFA ist es möglich, aus den Beobachtungen vieler sogenannter manifesten Variablen, beispielsweise Items eines Fragebogens, auf erklärende, latente Variablen, auch Faktoren genannt, zu schließen. Daher wird die EFA auch als ein datenreduzierendes Verfahren bezeichnet.

---

<sup>12</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 12

<sup>13</sup> Bortz, J. (2005): S. 561

<sup>14</sup> Bortz, J. (2005): S. 511

## Vorgehensweise: Variablenauswahl und Korrelationsmatrix

Das Fundament einer jeden EFA bildet der Korrelationskoeffizient. Dieser wird zur Konstruktion einer Korrelationsmatrix genutzt, so dass erste Bündelungen mehrerer Variablen zu erklärenden Faktoren möglich werden und aus der Matrix abgelesen werden können. Der Korrelationskoeffizient errechnet sich über die nachfolgende Formel:

$$r_{y_1, y_2} = \frac{\sum_{k=1}^K (y_{k1} - \bar{y}_1) \cdot (y_{k2} - \bar{y}_2)}{\sqrt{\sum_{k=1}^K (y_{k1} - \bar{y}_1)^2 \cdot \sum_{k=1}^K (y_{k2} - \bar{y}_2)^2}} \quad (3.1.)$$

Zur Vereinfachung der Analyse ist eine Standardisierung der Datenmatrix sinnvoll. Dadurch werden die Rechnungen einfacher, die Interpretation erleichtert und die Variablen vergleichbarer gemacht. Mit Hilfe der standardisierten Datenmatrix  $Z$  lässt sich die Korrelationsmatrix  $R$  gemäß der folgenden Formel berechnen:

$$R = \frac{1}{K-1} \cdot Z' \cdot Z \quad (3.2.)$$

Wie bereits erwähnt, lässt sich anhand der Korrelationsmatrix  $R$  erkennen, welche Variablen einen hohen Zusammenhang aufweisen und eventuell einen Faktor bilden können. Jedoch ist nicht ersichtlich, ob die Variablen kausal zusammenhängen oder durch eine weitere Variable erklärt werden. Um solche Fragestellungen im weiteren Verlauf der Faktorenanalyse beantworten zu können, muss die Korrelationsmatrix  $R$  eine gewisse Eignung aufweisen. Wenn nicht, ist eine sinnvolle Faktorenanalyse nicht möglich. Ein Hinderungsgrund für die Faktorenanalyse könnte zum Beispiel eine sehr heterogene Struktur der Erfassungsdaten sein. Dies würde sich durch viele kleine Werte in der Korrelationsmatrix ausdrücken. Deshalb sollte vor dem Durchführen einer Faktoranalyse getestet werden, ob die Daten normal- oder wenigstens gleichartig verteilt sind<sup>15</sup>. Einige geeignete Testverfahren dafür sind die Signifikanzprüfung der Korrelation, die Analyse der Inversen der Korrelationsmatrix, der Bartlett-Test, die Anti-Image-Kovarianz-Matrix und auch das Kaiser-Meyer-Olkin-Kriterium. In

---

<sup>15</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 339

meiner Arbeit werden für die Überprüfung der Dateneignung das Kaiser-Meyer-Olkin- und das MSA-Kriterium angewendet. Das KMO berechnet aufgrund einer Einteilung der Varianz der Variablen in abhängig und unabhängig den MSA, eine Prüfgröße, die als Indikator dafür dient, wie geeignet die vorliegenden Daten für eine Faktorenanalyse sind. Der Wert kann zwischen 0 und 1 liegen, wobei ein  $MSA > 0,5$  für die Analyse als erforderlich angesehen wird<sup>16</sup>. Da der MSA-Wert auch für einzelne Variablen berechnet werden kann, ist es möglich, durch einen iterativen Prüfungsprozess "schlechte" Variablen aus der Analyse herauszunehmen, um eine sinnvolle Anwendung der Faktoranalyse zu garantieren.

### **Extraktion der Faktoren und Bestimmung der Kommunalitäten**

Um von den Korrelationen zu den Faktorwerten zu gelangen, spielt das sogenannte Fundamentaltheorem der Faktoranalyse eine bedeutende Rolle. Damit dieses verstanden werden kann, müssen zunächst die dafür notwendigen Beziehungen bzw. Formeln hergeleitet werden:

Die Grundannahme der Faktoranalyse ist, dass sich „jeder Beobachtungswert einer Ausgangsvariablen  $x_j$  oder der standardisierten Variablen  $z_j$  [...] als eine Linearkombination mehrerer (hypothetischer) Faktoren beschreiben lässt“<sup>17</sup>:

$$z_{kj} = a_{j1} \cdot p_{k1} + a_{j2} \cdot p_{k2} + \dots + a_{jQ} \cdot p_{kQ} = \sum_{q=1}^Q a_{jq} \cdot p_{kq} \quad (3.3.)$$

Wobei die Beobachtungswerte durch die Faktoren  $p_{kq}$  und die jeweiligen Faktorladungen bei Merkmal  $j$  als  $a_{jq}$  beschrieben werden. Die Faktorladung gibt Auskunft darüber, wie viel Gewicht ein Faktor auf eine Ausgangsvariable hat. In anderen Worten, er spiegelt die Korrelation zwischen Faktor und Variable wieder.

Zur weiteren Vereinfachung wird die Formel (3.3.) häufig in Matrixschreibweise angegeben:

$$Z = P \cdot A' \quad (3.4.)$$

---

<sup>16</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 342

<sup>17</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 344

Aus dem vorherigen Abschnitt ist bekannt, dass sich die Korrelationsmatrix  $R$  mit der standardisierten Datenmatrix  $Z$  berechnen lässt:

$$R = \frac{1}{K-1} \cdot Z' \cdot Z \quad (3.2.),$$

sodass sich durch Einsetzen von 3.4. in 3.2. folgende Gleichung ergibt:

$$R = \frac{1}{K-1} \cdot (P \cdot A')' \cdot (P \cdot A') \quad (3.5.)$$

Durch umstellen kann man, da alle Werte standardisiert sind, folgenden Term zu einer Einheitsmatrix  $C$  zusammenfassen:

$$\frac{1}{K-1} \cdot P' \cdot P$$

Nach dieser Vereinfachung sieht die Formel wie folgt aus:

$$R = A \cdot C \cdot A' \quad (3.6.)$$

Mit dem Wissen um die Einheitsmatrix  $C$  kann die endgültige Gleichung folgendermaßen dargestellt werden:

$$R = A \cdot A' \quad (3.7.)$$

Die Gleichungen (3.6.) und (3.7.) werden als das Fundamentaltheorem der Faktoranalyse bezeichnet, da sie Informationen über den Zusammenhang zwischen Korrelations- und Faktorladungsmatrix geben. Dabei ist zu beachten, dass Gleichung (3.7.) nur gilt, wenn die Faktoren unabhängig voneinander sind.

Anhand einer Kosinustabelle lässt sich jeder Korrelation ein bestimmter Winkel zuordnen, so dass die Variablen als Vektoren dargestellt werden können. Durch orthogonale Transformationen erhält man voneinander unabhängige Faktoren, die senkrecht zu einander stehen. Die Winkel zwischen den Variablen und dem sich ergebenden Faktor, auch als Resultante bezeichnet, stellen die verschiedenen Faktorladungen dar und lassen sich wiederum über die Kosinustabelle ermitteln.

Aufgrund der standardisierten Ausgangs-Datenmatrix sollten, wenn je Variable ein Faktor entsteht, die aufsummierten Ladungsquadrate (= erklärte Varianz) 1 entsprechen. Tatsächlich ist allerdings mit einer erklärten Varianz kleiner als 1 zu rechnen.

Dies liegt in dem Verfahren der Faktorenanalyse als solches begründet, deren Ziel ja darin besteht, viele Variablen bei möglichst hoher Erklärung von Varianz auf wenige Faktoren zu reduzieren. Mit jedem zusätzlichen Faktor erhöht sich der Anteil an Varianz einer Variablen, der durch die Faktoren erklärt wird. Der Teil der Varianz, der durch die Faktoren erklärbar ist, wird Kommunalität  $h^2_j$  genannt. Daraus folgt, dass die Gleichung (3.7.), die mit Hilfe des Fundamentaltheorems aufgestellt wurde, um die spezifische Varianz  $U$  erweitert werden muss:

$$R = A \cdot A' + U \quad (3.8.)$$

Ein Hauptproblem bei der Faktorenanalyse ist, dass diese Kommunalitäten erst ermittelt bzw. geschätzt werden müssen. Im Laufe der Zeit wurden einige Verfahren entwickelt, um plausible Werte für die Kommunalitäten zu erhalten. Zu den wichtigsten zählen die Hauptkomponentenanalyse (PCA), die Hauptachsenanalyse (PAF), die Maximum-Likelihood-Faktorenanalyse (ML) und die Kleinste-Quadrate-Methode (KQ). Im Rahmen dieser Arbeit möchte ich genauer auf die Verfahren PCA und PAF eingehen.

Bei der PCA wird die Annahme getroffen, dass sich die Varianz aller Variablen vollständig durch die Extraktion der Faktoren erklären lässt, sodass alle Anfangswerte bei der Kommunalitätsschätzung auf "1" gesetzt werden. Ist die Anzahl der Faktoren geringer als die Anzahl der Variablen, folgt daraus im Umkehrschluss, dass ein bewusst in Kauf genommener Informationsverlust eintritt. Die PAF hingegen basiert auf der Annahme, dass sich die Varianz einer Variablen in die Kommunalität und in eine Einzelrestvarianz, die nicht erklärt werden kann, zerlegen lässt. Die Kommunalität wird bei der PAF entweder durch zusätzliche Informationen von dem Anwender festgelegt oder sie lässt sich über den Iterationsprozess der PAF ermitteln (mit Anfangswerten kleiner als 1). SPSS verwendet dazu beispielsweise als Startwert das multiple Bestimmtheitsmaß.

Die unterschiedlichen Ansätze der beiden Verfahren lassen sich zum Teil dadurch erklären, dass sie unterschiedliche Ziele verfolgen. Die PCA versucht, die Daten so gut wie möglich wiederzugeben, und zwar unter der Verwendung einiger weniger Faktoren. Im Gegensatz dazu will man mit der PAF den größtmöglichen Anteil an Varianz erklären, sodass es zwangsläufig notwendig ist, eine Unterscheidung in Kommunalität und Einzelrestvarianz vorzunehmen. Zusammengefasst heißt das, dass die

PCA nach einem Sammelbegriff für Faktoren mit hoch ladenden Variablen sucht und die PAF nach der Ursache für hohe Korrelationen<sup>18</sup>. Deshalb wird auch häufig die PCA als eigenes Verfahren neben der Faktorenanalyse betrachtet, so dass man im eigentlichen Sinne nur bei der PAF von einer Faktorenanalyse sprechen dürfte.

### Zahl der Faktoren und Faktorinterpretation

Es gibt keine Verfahren, die bestimmen, wie Faktoren festzulegen sind bzw. wie viele gewählt werden sollten. Allerdings gibt es auch an dieser Stelle wieder statistische Methoden, die bei der Entscheidungsfindung beachtet werden sollten. Hierbei sind besonders das Kaiser-(Eigenwert-)Kriterium und der Screen-Test von Bedeutung. Zum Verständnis des Kaiser-Kriteriums muss zunächst geklärt werden, welche Bedeutung die Bezeichnung "Eigenwert" im Zusammenhang mit einer Faktoranalyse hat. Addiert man in einer Faktorladungsmatrix zeilenweise die quadrierten Faktorladungen auf über die verschiedenen Faktoren, so erhält man die Kommunalität der Variablen. Summiert man dagegen die quadrierten Faktorladungen spaltenweise, also über alle Variablen auf, so erhält man den Eigenwert des jeweiligen Faktors. Nun besagt das Kaiser-Kriterium, dass man nur solche Faktoren extrahieren sollte, deren Eigenwert über 1 liegt.

Faktorenmatrix					Kommunalität
Variablen	Faktor 1		Faktor 2		
	Faktorl. 1	(Faktorl.1)^2	Faktorl. 2	(Faktorl.2)^2	
Variable 1	0,81	0,6561	0,54	0,2916	<b>0,9477</b>
Variable 2	0,85	0,7225	-0,32	0,1024	<b>0,8249</b>
Variable 3	-0,18	0,0324	0,84	0,7056	<b>0,738</b>
Variable 4	0,15	0,0225	0,91	0,8281	<b>0,8506</b>
Eigenwert		<b>1,4335</b>		<b>1,9277</b>	

Abbildung 3.1.: Beispielmatrix mit Eigenwerten und Kommunalitäten

In der abgebildeten Matrix - Abbildung 3.1. - wird dieser Sachverhalt beispielhaft dargestellt. In der zweiten und vierten Spalte stehen die Faktorladungen zweier Faktoren bezüglich der jeweiligen Variablen. In Spalte 3 und 5 stehen die quadrierten Faktorladungen der zwei Beispielfaktoren. Die Kommunalitäten ergeben sich, wenn man

<sup>18</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 357

die quadrierten Faktorladungen zeilenweise addiert. Die Eigenwerte errechnen sich durch spaltenweises Addieren der quadrierten Faktorladungen. In dem dargestellten Modell werden 94,77% der Varianz von Variable 1 durch die zwei Faktoren erklärt (= Kommunalität). Dies wäre ein hoch einzustufender Wert. Die Variablen 1 und 2 laden hoch auf den Faktor 1 (0,81 und 0,85), sodass man diese beiden Variablen durch Faktor 1 zusammenfassen könnte. In entsprechender Weise gilt das gleiche für die Variablen 3 und 4 bezüglich des Faktors 2 (0,84 und 0,91). Nach dem Kaiser-Kriterium würde man beide Faktoren extrahieren, da für beide die Eigenwerte über 1 liegen. Der "kritische" Wert ist genau 1, da von einem Faktor erwartet wird, dass er mindestens die Varianz einer einzelnen Variable erklärt und genau diese ist bei einer standardisierten Variable gleich 1.

Der Screen-Test stellt eine grafische Lösung für die Ermittlung der Anzahl an Faktoren dar. Dabei werden die Eigenwerte in einem Koordinatensystem eingetragen, beginnend mit dem höchsten.

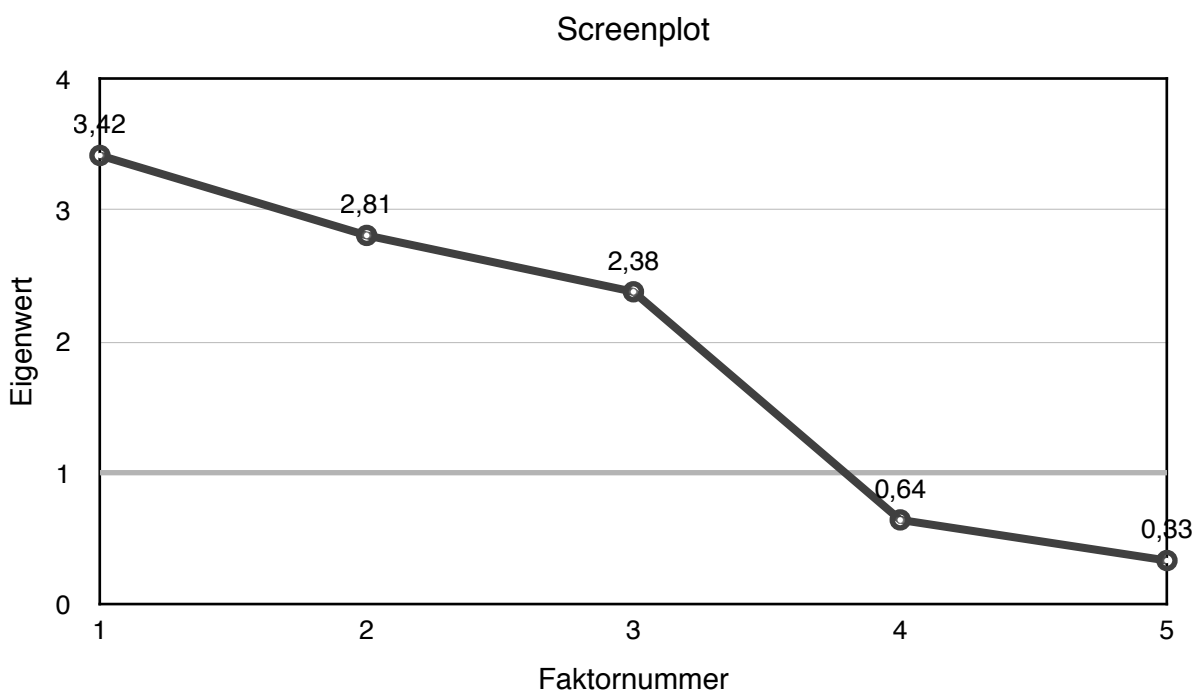


Abbildung 3.2.: Screen-Test Beispiel

Wie der Abbildung 3.2. entnommen werden kann, wurden 5 (ausgedachte) Eigenwerte in einem Koordinatensystem dargestellt. Zusätzlich ist bei dem Eigenwert 1 das Kaiser-Kriterium durch eine Gerade über alle Faktoren eingezeichnet. Der Screen-Test besagt nun, dass der Eigenwert, der links des größten Gefälles, in diesem Zusammenhang auch als Knick bezeichnet, liegt, die Anzahl der Eigenwerte bestimmt.



Eine entsprechende Auswertung würde ergeben, dass zwischen dem dritten und vierten Faktor mit 1,74 die größte Differenz vorliegt. Demnach wären in diesem Beispiel drei zu extrahierende Faktoren auszuwählen. Allerdings muss der Screen-Test nicht immer ein zuverlässiges Mittel zur Feststellung der Anzahl an Faktoren sein, denn bei ähnlichen Differenzen ist nicht immer ein eindeutiger Knick zu erkennen. Das Kaiser-Kriterium findet deshalb auch häufiger Anwendung.

Zur Interpretation der Faktoren ist es hilfreich, wenn die Faktorladungsmatrix in der Einfachstruktur vorliegt. Das heißt, dass jede Variable nur hoch auf jeweils einen Faktor lädt. Dadurch ist einfacher zu bestimmen, welche Variablen den Faktor "ausmachen". Bei empirischen Analysen, vor allem mit vielen Variablen sind derartige Einfachstrukturen nicht realistisch. Um trotzdem möglichst einfach und einheitliche Prozesse für die Interpretation der Faktoren an der Hand zu haben, wurden bestimmte Regeln aufgestellt und Verfahren vereinfacht. Beispielsweise wird angenommen, dass eine Ladung hoch ist, wenn sie größer als 0,5 ist<sup>19</sup>. Dies stellt allerdings ein Problem dar, wenn eine Variable für mehrere Faktoren eine Ladung größer als 0,5 besitzt, da die Variable dann für beide Faktoren eine relevante Rolle spielt. Um dieses Problem zu lösen, kann eine Rotation des Koordinatenkreuzes im Ursprung durchgeführt werden. Bei der Rotation unterscheidet man zwei verschiedene Verfahren. Es gibt die orthogonale Rotation, die der Annahme unterliegt, dass die Faktoren nicht korreliert sind, sodass bei der Durchführung die Faktorenachsen weiterhin rechtwinklig zueinander stehen. Besteht die Annahme, dass die Faktoren korreliert sind, werden die Faktorenachsen schief gedreht. Dieses Verfahren wird oblique oder schiefwinklige Rotation genannt. Die Rotationen dienen als Interpretationshilfe, denn durch das Einsetzen von Rotationen kann erreicht werden, dass eine Variable, die anfangs auf mehrere Faktoren lädt, nur noch auf einen Faktor hoch geladen ist.

### **Bestimmung der Faktorwerte**

Die Faktorwerte dienen zum Einen zur Klärung, welche Variablen sich zu einem gemeinsamen Faktor zusammenfassen lassen. Zum Anderen dienen sie auch der Feststellung "welche Werte die Objekte nun hinsichtlich der extrahierten Faktoren

---

<sup>19</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 362

annehmen”<sup>20</sup>. Dazu benötigt man neben den Faktoren selbst auch deren Ausprägungen bezüglich der Objekte. Aus Gleichung 3.4. ist bekannt:

$$Z = P \cdot A' \quad (3.4.)$$

$Z$  ist als standardisierte Datenmatrix bekannt und die Faktorladungen  $A$  lassen sich wie oben beschrieben berechnen. Das bedeutet, dass diese Gleichung nach  $P$  aufgelöst werden muss, um die Faktorwerte zu erhalten. Somit ergibt sich die Gleichung:

$$P = Z \cdot (A')^{-1} \quad (3.9.)$$

Es kann allerdings passieren, dass sich von der Faktorladungsmatrix  $A$  keine Inverse bilden lässt. Nämlich genau dann, wenn sie aufgrund der gewollten Variablenreduktion nicht quadratisch ist. Eine quadratische invertierbare Matrix wird in einem solchen Fall erzeugt, indem von rechts mit der Matrix  $A$  multipliziert wird. Daraus ergibt sich die Formel 3.10.:

$$P = Z \cdot A \cdot (A' \cdot A)^{-1} \quad (3.10.)$$

Alternativ können die Faktorwerte auch über eine Regression geschätzt werden. Dabei wird die standardisierte Ausgangsmatrix  $Z$  mit dem Regressionkoeffizienten multipliziert. Dadurch können die Faktorwerte quasi “als Ergebnis” der Faktoranalyse grafisch abgebildet werden. Für eine Extraktion von 2 oder 3 Faktoren ist dies in einem Koordinatensystem besonders gut möglich. Auf diese Weise lässt sich optisch nachvollziehen, wie stark die Variablen und die Faktoren (= die Achsen) zusammenhängen.

### 3.2.1.2. Clusteranalyse

Während die Faktorenanalyse das Ziel hat, die Variablen zu bündeln, sollen mit Hilfe einer Clusteranalyse Objekte zu sogenannten Clustern gebündelt werden, sodass man homogene Gruppen erhält, die jedoch untereinander eine Heterogenität aufweisen. Analog zur Faktoranalyse bezeichnet der Begriff Clusteranalyse eine Gruppe

---

<sup>20</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 364

von Analyseverfahren. Sie unterscheiden sich vor allem in zwei Gesichtspunkten: Zum Einen in dem jeweiligen Proximitätsmaß, dem Maß, mit dem die Ähnlich- bzw. Unähnlichkeit zwischen den Objekten ermittelt wird, und zum Anderen in dem angewandten Gruppierungsverfahren. Dieses legt fest, auf welche Weise die Objekte zu Gruppen zusammengefasst werden. Dabei unterscheidet man in Fusionierungsalgorithmen und Partitionierungsalgorithmen. Weiterhin wichtig ist, dass bei der Clusteranalyse nicht nur einige sondern alle erhobenen Variablen in die Analyse mit einbezogen werden.

Eine Clusteranalyse kann grundlegend in 3 Schritte unterteilt werden<sup>21</sup>. Zunächst müssen die Ähnlichkeiten bzw. Distanzen zwischen den Beobachtungen bestimmt werden. Anschließend wählt der Anwender einen Fusionierungsalgorithmus aus. Zum Schluss muss noch die Anzahl der Cluster festgelegt werden. Da mit dem Begriff Clusteranalyse diverse Verfahren gemeint sind, werde ich auch im Folgenden mehrere verschiedene Varianten vorstellen.

### **Bestimmung der Ähnlichkeiten bzw. Distanzen**

Am Anfang einer jeden Clusteranalyse steht eine Rohdatenmatrix. Diese beinhaltet für jedes Objekt (z. B. ein Teilnehmer einer Umfrage) die Variablenwerte (Antworten auf die Fragen). Wie eine solche Matrix aufgebaut sein kann, ist in Abbildung 3.3. dargestellt.

	Variable 1	Variable 2	...	Variable J
Objekt 1				
Objekt 2				
...				
Objekt K				

*Abbildung 3.3.: Muster einer Rohdatenmatrix*

Um Ähnlichkeiten oder Unterschiede erkennen zu können, wird die Rohdatenmatrix in eine quadratische Distanz- oder Ähnlichkeitmatrix umgewandelt. Dabei werden die einzelnen Objekte in den Reihen und Spalte gemäß Abbildung 3.4. gegenübergestellt.

<sup>21</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 398

	Objekt 1	Objekt 2	...	Objekt K
Objekt 1				
Objekt 2				
...				
Objekt K				

*Abbildung 3.4.: Muster einer Distanz- oder Ähnlichkeitsmatrix*

Um in dieser Matrix die vorliegenden Ähnlich- bzw. Unähnlichkeiten (Distanzen) der jeweiligen Variablenwerte zwischen den untersuchten Objekten darzustellen, können verschiedene Proximitätsmaße verwendet werden. Die Ermittlung der Werte für eine Distanz- oder Unähnlichkeitsmatrix sind der folgenden Beschreibung und der dazu gehörigen Abbildung 3.5. zu entnehmen. Prinzipiell gilt, je höher der ermittelte Wert eines Distanzmaßes (Wertebereich zwischen 0 und 1) ist, desto unähnlicher sind sich die Objekte; je tiefer der Wert, desto ähnlicher sind sich die Objekte. Für Ähnlichkeitsmaße gilt der umgekehrte Zusammenhang, so dass ein Wert von "1" für komplette Übereinstimmung aller Variablen bei zwei Objekten steht.

Die anzuwendenden Proximitätsmaße hängen jedoch stark von der Skalierung der vorliegen Daten ab. Im Rahmen dieser Arbeit werde ich hauptsächlich Verfahren für binäre und für metrische Variablen vorstellen.

Binäre Variablen nehmen entweder den Wert 0 - Eigenschaft trifft nicht zu - oder den Wert 1 - Eigenschaft trifft zu - an. Die Clusteranalyse führt jeweils einen paarweisen Vergleich aller Eigenschaftswerte von zwei Objekten durch. Das heißt, bei jedem Vergleich kann man gemäß Abbildung 3.5. vier verschiedene Fälle unterscheiden.

	Objekt 2	Eigenschaft vorhanden (1)	Eigenschaft nicht vorhanden (0)	Zeilensumme
Objekt 1				
Eigenschaft vorhanden (1)		a	c	a + c
Eigenschaft nicht vorhanden (0)		b	d	b + d
Spaltensumme		a + b	c + d	M

Abbildung 3.5.: Paarweiser Vergleich von binären Variablen

Als Grundansatz für die Berechnung von Ähnlichkeiten bei binären Variablen lässt sich die folgende Formel verwenden:

$$S_{ij} = \frac{a + \delta \cdot d}{a + \delta \cdot d + \lambda \cdot (b + c)} \quad (3.11.)$$

Wobei  $S_{ij}$  für die Ähnlichkeit und  $\delta$ ,  $\lambda$  für mögliche (konstante) Gewichtungsfaktoren stehen. Je nachdem, welchen Koeffizienten man betrachtet, unterscheiden sich die Gewichtungsfaktoren:

Wählt man den Jaccard-Koeffizienten mit Delta gleich "0" und Lambda gleich "1", ergibt sich eine Formel, bei der  $a$  (die Anzahl der Eigenschaften, die bei beiden Objekten vorhanden sind) durch die Summe aus  $a$  und  $b$  und  $c$  ( $b$  und  $c$  sind jeweils die Anzahl der Eigenschaften, die nur bei einem Objekt vorhanden sind) geteilt wird. Eigenschaften, die bei beiden Objekten nicht vorliegen, werden von dem Jaccard-Koeffizienten nicht berücksichtigt.

Im Gegensatz dazu erfolgt dies bei dem Koeffizient von Russel und Rao (RR-Koeffizient). Bei dessen Berechnung steht im Zähler auch  $a$ , jedoch im Nenner die Summe aus  $a$ ,  $b$ ,  $c$  und  $d$ , was sich nach der Tabelle aus Abbildung 3.5. mit  $M$  zusammenfassen lässt. Der RR-Koeffizient nimmt den gleichen Wert an wie der Jaccard-Koeffizient, solange keine Fälle vorliegen, in denen beide Objekte, die Eigenschaft nicht aufweisen ( $d$ ), da dadurch der Nenner größer wird und das Ähnlichkeitsmaß sinkt.

Bei dem Simple-Matching- oder M-Koeffizienten wird zusätzlich auch noch das paarweise Nichtvorliegen als Übereinstimmung gewertet. Das spiegelt sich darin wieder, dass der Zähler aus der Summe von  $a$  und  $d$  besteht, welcher wie bei dem RR-Koeffizienten

fizienten durch  $M$  geteilt wird. Prinzipiell sind die Werte des M-Koeffizienten höher als die des RR- und des Jaccard-Koeffizienten.

Für alle Koeffizienten lässt sich nach Ermittlung aller Ähnlichkeitswerte eine symmetrische Ähnlichkeitsmatrix aufstellen, die die Ähnlichkeitsmaße aller Objekte untereinander angibt. Alle Koeffizienten haben gemeinsam, dass sie zum selben Ergebnis kommen, solange keine Eigenschaft bei beiden Objekten nicht vorliegt, also  $d$  gleich "0" ist. Welcher der drei Koeffizienten bei der Analyse angewendet werden sollte, lässt sich nicht allgemeingültig festlegen. Eine entscheidende Rolle bei der Koeffizientenauswahl spielt allerdings, ob das Nichtvorhandensein einer Eigenschaft die gleiche Aussagekraft hat wie das Vorhandensein. Wenn das Vorhandensein einer Eigenschaft gleichermaßen bedeutsam ist wie das Nichtvorhandensein, so ist der M-Koeffizient vorzuziehen. Andernfalls sollte man als Proximitätsmaß den Jaccard-Koeffizienten oder einen verwandten Schätzer wählen.

Die Analyse von Daten mit nominaler Skalierung soll an dieser Stelle nur kurz erwähnt werden, da sie in der vorliegenden Arbeit keine Anwendung findet. Ein nominales Merkmal, das mehr als zwei Ausprägungen besitzt, kann durch eine Datentransformation in eine "binäre (Hilfs-)Variable"<sup>22</sup> überführt werden, sodass die zuvor beschriebenen Proximitätsmaße für binäre Modelle zum Einsatz kommen können. Bei Häufigkeitsvariablen lässt sich über den Chi-Quadrat-Homogenitätstest untersuchen, ob den jeweiligen Grundgesamtheiten die gleichen Verteilungen zu Grunde liegen. Wird die Nullhypothese gleicher Verteilungen abgelehnt, könnte man annehmen, dass die Grundgesamtheiten unterschiedlich sind. Mit Hilfe des Chi-Quadrat-Maßes, das auf dem gleichnamigen Test basiert, lässt sich eine Distanzmatrix aufstellen, die im Gegensatz zu den Ähnlichkeitsmatrizen für binäre Variablen als Unähnlichkeitsmatrix bezeichnet wird.

Für metrisch skalierte Variablen sollen an dieser Stelle die Minkowski-Metrik, die einfache und quadrierte Euklidische Distanz und der Q-Korrelationskoeffizient vorgestellt werden. Die Formel für die Minkowski-Metrik sieht wie folgt aus:

$$d_{k,l} = \left[ \sum_{j=1}^J |x_{kj} - x_{lj}|^r \right]^{\frac{1}{r}} \quad (3.12.)$$

---

<sup>22</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 407

Dabei steht  $d_{k,l}$  für die Distanz der Objekte k und l.  $x_{kj}$  und  $x_{kl}$  sind die Werte der Variablen j bei den Objekten k und l.  $r \geq 1$  stellt die Minkowski-Konstante dar. Wird beispielsweise  $r = 1$  gewählt, so spricht man von der L1-Norm oder auch von der City-Block-Metrik. In diesem Falle summiert die Formel die Differenzen zwischen den beiden verglichenen Eigenschaften über alle Eigenschaften auf, womit ein Maß für die Unähnlichkeit der beiden Objekte entsteht. Wird  $r = 2$  gewählt, spricht man von der (einfachen) Euklidischen Distanz, die auch L2-Norm genannt wird. Bei der quadrierten Euklidischen Distanz werden die einzelnen Differenzen der jeweiligen Eigenschaften quadriert, jedoch wird anschließend nicht aus der Summe die Wurzel gezogen. D.h., die einfache Euklidische Distanz ist nichts anderes als die Wurzel aus der quadrierten Euklidischen Distanz. Der Unterschied in der Aussage der beiden Verfahren besteht darin, dass bei der quadrierten Version große Unterschiede in den Eigenschaften stärker gewichtet werden.

Zur Analyse metrischer Daten mit Hilfe eines Ähnlichkeitsmaßes kann aber auch der Q-Korrelationskoeffizient, auch Pearson-Korrelationskoeffizient genannt, genutzt werden.

$$r_{k,l} = \frac{\sum_{j=1}^J (y_{jk} - \bar{y}_k) \cdot (\bar{y}_{jl} - \bar{y}_l)}{\left\{ \sum_{j=1}^J (y_{jk} - \bar{y}_k)^2 \cdot \sum_{j=1}^J (\bar{y}_{jl} - \bar{y}_l)^2 \right\}^{\frac{1}{2}}} \quad (3.13.)$$

In der Formel bezeichnet  $y_{jk}$  die Ausprägung der Eigenschaft j bei Objekt k mit  $j = 1, 2, \dots, J$  und das  $\bar{y}_k$  den Durchschnittswert aller Eigenschaften bei Objekt k. Als Ergebnis einer paarweisen Berechnung des Koeffizienten ergibt sich eine Ähnlichkeitsmatrix. Ob man für eine metrische Variable nun ein Distanz- oder Ähnlichkeitsmaß verwendet, hängt sehr von inhaltlichen Überlegungen ab. Man sollte mit Distanzmaßen arbeiten, wenn die absolute Distanz zwischen Objekten als wichtig angesehen wird und große Unterschiede ebenfalls von großer Bedeutung sind. Ähnlichkeitsmaße sind zu bevorzugen, wenn auf grundsätzliche Ähnlichkeit geachtet wird.

## Auswahl des Fusionsalgorithmus

Anhand der über das Proximitätsmaß berechneten Distanz- oder Ähnlichkeitsmatrizen lassen sich nun mit Hilfe von Clusteralgorithmen mehrere Objekte zu Gruppen zusammenschließen. Dabei kann eine Einteilung der Verfahren durch bestimmte Kriterien vorgenommen werden. Wird von dem Algorithmus nur eine einzige Variable zur Gruppierung betrachtet, spricht man von monothetischen Verfahren. Polythetischen Verfahren betrachten alle Variablen gleichzeitig. Sie spielen bei der Clusteranalyse eine weitaus wichtigere Rolle und werden deshalb auch in dieser Arbeit angewendet. Des Weiteren findet eine Unterscheidung bei den Fusionierungsprozessen statt. Es gibt graphentheoretische Verfahren, partitionierende Verfahren, hierarchische Verfahren und Optimierungsverfahren. Aufgrund der größeren Bedeutung für empirisches Arbeiten, werden im Rahmen dieser Arbeit die partitionierenden und hierarchischen Verfahren näher betrachten.

Der Grundgedanke der partitionierenden Verfahren ist, dass der Algorithmus bereits mit zufällig in Gruppen eingeteilten Objekten beginnt. Damit einhergehend ist auch schon zu Beginn die Anzahl der Cluster festgelegt. Dabei werden die Objekte durch den Algorithmus solange verschoben, bis eine Zielfunktion optimiert ist. Dieses Austauschverfahren kann in 6 Schritten zusammengefasst werden:

1. Eine Anfangspartition wird vorgegeben.
2. Pro Gruppe und Eigenschaft wird das arithmetische Mittel berechnet.
3. Für die vorliegende Gruppenverteilung wird die Fehlerquadratsumme für alle Gruppen berechnet.
4. Es wird untersucht, ob durch Austausch von Objekten das Varianzkriterium<sup>23</sup> verringert werden kann.
5. Das Objekt, das die Varianz am meisten verringert, wird in die jeweilige Gruppe verschoben.
6. Neue Mittelwertberechnung für die beiden betroffenen Gruppen.

---

<sup>23</sup> Weitere Erläuterungen zu dem Varianzkriterium erfolgen bei den hierarchischen Verfahren.



Das Verfahren beginnt von Neuem bei Schritt Nummer drei und ist beendet, wenn durch eine Verschiebung eines Objektes keine Varianzminimierung mehr möglich ist. Häufig werden dabei allerdings nur lokale und nicht globale Optima erreicht, da aufgrund der Vielzahl von Kombinationsmöglichkeiten keine vollständige Enumeration durchgeführt werden kann. Daher ist es möglich durch eine Veränderung der Starteinteilung ein besseres Ergebnis zu erhalten, was sich mit dem Varianzkriterium überprüfen lassen kann. Zu Beginn einer partitionierenden Clusteranalyse muss der Anwender zwei Entscheidungen treffen. Zum einen muss er wählen, in wie viele Gruppen die Einteilung stattfinden soll und zum Anderen muss festgelegt werden, wie die Startgruppen eingeteilt werden sollen. Entweder durch eine Zufallszuweisung, eine Durchnummerierung oder das Zurückgreifen auf hierarchische Verfahren.

Bei den hierarchischen Verfahren wird in agglomerative und divisive Verfahren unterteilt. Während agglomerative Verfahren zu Beginn in die kleinste mögliche Partition einteilen (je Objekt eine Gruppe), starten die divisiven Verfahren mit der größtmöglichen Partition (alle Objekte in einer Gruppe). Aufgrund der höheren Bedeutung werde ich mich an dieser Stelle auf die agglomerativen Verfahren beschränken, die sich ebenfalls in 6 allgemeine Schritte aufteilen lassen:

1. Einteilung in die feinste mögliche Partition. (Je Objekt eine Gruppe)
2. Für alle in der Analyse eingeschlossenen Objekte werden die Ähnlichkeits- bzw. Distanzmaße berechnet.
3. Die Cluster mit der größten Ähnlichkeit oder geringsten Distanz werden gesucht.
4. Die Cluster mit der größten Ähnlichkeit oder geringsten Distanz werden zu einer Gruppe zusammengefasst. Zum Ende dieses Schrittes liegt ein Cluster weniger vor als zu Beginn.
5. Es wird die Ähnlichkeit bzw. Distanz zwischen den neuen Gruppen berechnet, so dass man eine reduzierte Matrix erhält.
6. Das Vorgehen von Schritt 3 bis 5 wird so lange wiederholt, bis alle Objekte einer Gruppe zugeteilt wurden.

Einige der agglomerativen Verfahren sind Single-Linkage, Complete-Linkage, Average-Linkage, Centroid, Median und Ward. Analog zu der Berechnung von Ähnlichkei-

ten, kann an dieser Stelle eine allgemeine Formel für das Ermitteln von Distanzen aufgestellt werden:

$$D(R;P+Q) = A \cdot D(R,P) + B \cdot D(R,Q) + E \cdot D(P,Q) + G \cdot |D(R,P) - D(R,Q)| \quad (3.14.)$$

mit  $D(R;P+Q)$  = Distanz zwischen R und der neuen Gruppe (P+Q)

mit  $D(R,P)$  = Distanz zwischen den Gruppen R und P

mit  $D(R,Q)$  = Distanz zwischen den Gruppen R und Q

mit  $D(P,Q)$  = Distanz zwischen den Gruppen P und Q

A, B, E und G stehen für die Konstanten, die bei jedem der 6 genannten Verfahren anders bestimmt sind. Wie genau, ist der folgenden Tabelle (Abbildung 3.6.) zu entnehmen.

Verfahren	A	B	C	G
Single-Linkage	0,5	0,5	0	-0,5
Complete-Linkage	0,5	0,5	0	0,5
Average-Linkage (ungewichtet)	0,5	0,5	0	0
Average-Linkage (gewichtet)	$NP/(NP+NQ)$	$NP/(NP+NQ)$	0	0
Centroid	$NP/(NP+NQ)$	$NP/(NP+NQ)$	$-(NP \cdot NQ)/(NP+NQ)^2$	0
Median	0,5	0,5	-0,25	0
Ward	$(NR+NP)/(NR+NP+NQ)$	$(NR+NP)/(NR+NP+NQ)$	$-NR/(NR+NP+NQ)$	0

Abbildung 3.6.: Berechnung der Distanzen bei hierarchischen, agglomerativen Verfahren

In der Tabelle sind  $NR$ ,  $NP$  und  $NQ$  die Anzahl der Objekte in der jeweiligen Gruppe. Hinsichtlich der Auswahl der Verfahren ist zu bemerken, dass die ersten vier Verfahren in der Tabelle mit jedem Proximitätsmaß arbeiten können. Bei den letzten drei Verfahren sollte hingegen ein Distanzmaß eingesetzt werden. An dieser Stelle möchte ich die Verfahren Single-Linkage, Complete-Linkage und Ward genauer vorstellen.

Wendet man die Single-Linkage-Werte aus Abbildung 3.6. auf die Formel 3.14. an, so erhält man die Formel:

$$D(P;P+Q) = 0,5 \cdot \{D(R,P) + D(R,Q) - |D(R,P) - D(R,Q)|\} \quad (3.15.)$$

Diese lässt sich auch vereinfachen zu:

$$D(R,P+Q) = \min\{D(R,P);D(R,Q)\} \quad (3.16.)$$

Sie besagt, dass bei der Zusammenfassung von P und Q zu einem Cluster die neue Distanz zwischen P+Q und R das Minimum von P zu R und Q zu R ist. Sind zum Beispiel die Distanzen zwischen R und P 5, zwischen Q und R 10 und zwischen P und Q 1, so würde man P und Q zusammenfassen, da sie die geringste Distanz haben. Die neue Gruppe P+Q hätte zu R eine Distanz von 5, da dieser Wert das vorliegende Distanzminimum zu R ist. Anschließend liegt eine vereinfachte, um ein Objekt reduzierte Distanzmatrix vor und das Prozedere beginnt von Neuem. Und zwar solange, bis alle Elemente zugeordnet worden sind. Dabei besitzt die Single-Linkage-Methode den Vorteil, dass Ausreißer erkannt werden können, da stets der kleinere Wert als neue Distanz festgelegt wird.

Für das Complete-Linkage Verfahren ergibt sich aus der Formel 3.14.:

$$D(P;P+Q) = 0,5 \cdot \{D(R,P) + D(R,Q) + |D(R,P) - D(R,Q)|\} \quad (3.17.),$$

Auch hier ist eine Vereinfachung möglich zu der Formel:

$$D(R,P+Q) = \max\{D(R,P);D(R,Q)\} \quad (3.18.)$$

Die Formel ist analog zu der der Single-Linkage-Methode zu benutzen, nur dass die maximale Entfernung als neue Distanz gewählt wird. Bei der Nutzung dieses Verfahrens werden eher kleinere Gruppen gebildet. Außerdem ist es nicht dazu geeignet, Ausreißer zu identifizieren.

Das Ward-Verfahren unterscheidet sich hinsichtlich der Fusionsart und der Distanzbildung deutlich von den beiden Linkage-Verfahren. Das Ward-Verfahren nutzt die folgende Formel, um die Differenz zwischen den neuen Gruppen zu berechnen:

$$D(R;P+Q) = \frac{1}{NR+NP+NQ} \cdot \{(NR+NP) \cdot D(R,P) + (NR+NQ) \cdot D(R,Q) - NR \cdot D(P,Q)\} \quad (3.19)$$

Der größte Unterschied in der Vorgehensweise bei der Fusion ist, dass nicht die Objekte mit der geringsten Distanz zusammengefasst werden, sondern jene, bei denen das Homogenitätsmaß am geringsten bleibt. Das heißt, es werden die Objekte zusammengefasst, die die Varianz am wenigsten erhöhen. Als Maß dafür wird das Varianzkriterium, auf das bereits im vorangegangenen Text verwiesen wurde, herangezogen. Es wird auch als Fehlerquadratsumme bezeichnet:

$$V_g = \sum_{k=1}^{K_g} \sum_{j=1}^J (y_{kjg} - \bar{y}_{jg})^2 \quad (3.20.)$$

Dabei stehen  $V_g$  für die Fehlerquadratsumme der jeweiligen Gruppe,  $y_{kjg}$  für den Beobachtungswert der Variablen  $j$  ( $j=1, 2, \dots, J$ ) bei Objekt  $k$  ( $k=1, 2, \dots, K_g$ ) in der jeweiligen Gruppe und  $\bar{y}_{jg}$  für den Mittelwert der Beobachtungswerte der Variablen  $j$  in der entsprechenden Gruppe. So versucht das Ward-Verfahren "diejenigen Objekte (Gruppen), die die Fehlerquadratsumme am wenigsten erhöhen"<sup>24</sup>, zusammenzufassen. Dieser Prozess wird ebenfalls so lange wiederholt, bis alle Objekte zu Gruppen zusammengeführt worden sind.

Es ist empfehlenswert, die verschiedenen Verfahren zu kombinieren und die Vorteile eines jeden Verfahrens auszunutzen. Beispielsweise kann man zunächst mit Hilfe des Single-Linkage-Verfahrens den Datensatz auf Ausreißer untersuchen und anschließend eine Analyse mit dem Ward-Verfahren durchführen, da es bei richtiger Anwendung, ein "sehr gutes"<sup>25</sup> Fusionierungsverfahren ist.

### **Bestimmung der Clusterzahl**

Bisher ist bekannt, wie die Beobachtungen der einzelnen Objekte hinsichtlich ihrer Distanzen und Ähnlichkeiten untersucht und wie die Objekte in Cluster zusammengefügt werden können. Ausführlich beschrieben wurden ferner die agglomerativen Verfahren, bei denen der Algorithmus mit der kleinstmöglichen Partition (je Objekt eine Gruppe) startet und welcher so lange ausgeführt wird, bis eine große Gruppe entstanden ist. Bestehen bleibt allerdings noch die Frage, welche Anzahl von Clustern die beste ist. Dieses Thema soll in dem folgenden Abschnitt behandelt werden.

---

<sup>24</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 427

<sup>25</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 431

Eine Methode zur Festlegung der Clusterzahl ist das Elbow-Kriterium. Hier soll in den Daten ein Sprung in der Änderung des Varianzkriteriums erkannt werden. Noch besser ersichtlich ist das Elbow-Kriterium in einem Screen-Plot, wenn man die Fehlerquadratsumme auf der y- und die Clusteranzahl auf der x-Achse abträgt. Im Idealfall ergibt sich eine Kurve, die einer Wurzelfunktion ähnelt. Liegt eine derartige Struktur vor, sollte man den Wert als Clusteranzahl wählen, der sich auf dem Graphen als "Ellenbogen" identifizieren lässt.

Da die Einschätzung des Elbow-Kriteriums sehr von der individuellen Betrachtung des Anwender abhängt, gibt es eine Reihe von statistischen Verfahren, die bei der Festlegung der Clusteranzahl unterstützen können, so dass objektivere Ergebnisse möglich werden. Eines davon ist das Calinski-Harabasz-Kriterium. Das Kriterium setzt metrisch-skalierte Variablen voraus und analysiert die Streuung innerhalb einer Gruppe (W), sowie zwischen den einzelnen Gruppen (B). Für verschiedene Cluster-Modelle mit jeweils K Gruppen wird die CH-Statistik ermittelt, mit  $\bar{y}_k$  als Mittelwert der entsprechenden Gruppe und  $\bar{y}$  als Mittelwert aller Beobachtungen.

$$CH(K) = \frac{tr(B)}{tr(W)} \frac{n-K}{K-1} \quad (3.21.)$$

$$B = \sum_{k=1}^K n_k (\bar{y}_k - \bar{y}) \cdot (\bar{y}_k - \bar{y})' \quad W = \sum_{k=1}^K \sum_{j=1}^{n_k} n_k (\bar{y}_{kj} - \bar{y}_k) \cdot (\bar{y}_{kj} - \bar{y}_k)'$$

Hat die CH-Statistik ein Maximum von K, zwischen 2 und (N-1) Gruppen, so kann man annehmen, dass in den Daten eine Clusterung in K-Gruppen vorliegt. Erhält man einen konstant fallenden Wert bei steigender Clusterzahl, kann vermutet werden, dass der Datensatz keine Struktur besitzt.

Als letztes weiteres Hilfsmittel zur Bestimmung der Clusteranzahl soll der Test von Mojena<sup>26</sup> eingeführt werden. Der dafür betrachtete Wert ist der standardisierte Fusionskoeffizient  $A_i$ , der sich je Fusionsstufe wie folgt berechnet:

---

<sup>26</sup> Der Test von Mojena ist vielmehr eine Kennzahl als ein statistischer Test.

$$A_i = \frac{\alpha_i - \bar{\alpha}}{s_{\alpha}} \quad (3.22.)$$

mit:

$$\bar{\alpha} = \frac{1}{n-1} \sum_{i=1}^{n-1} \alpha_i \quad s_{\alpha} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n-1} (\alpha_i - \bar{\alpha})^2}$$

Die Faustregel für diesen Test lautet: Man wählt diejenige Gruppenszahl als Lösung, bei der ein bestimmter Wert der Teststatistik das erste Mal überschritten wird. Bei der Wahl dieses Wertes gibt es verschiedene Ansichten. Mojena selbst schlägt einen Wert von 2,75 vor; andere<sup>27</sup> empfehlen abhängig von der Datenstruktur einen Wert zwischen 1 und 2.

### 3.2.2. Anwendung auf den vorliegenden Datensatz

In diesem Abschnitt sollen die zuvor beschriebenen Verfahren der Faktoren- und Clusteranalyse auf den vorliegenden Datensatz angewendet werden. Die Faktorenanalyse dient dazu, in der Stichprobe vorhandene Strukturen bzw. Faktoren, die die Berufsvorbereitung der Schüler bestimmen, aufzudecken. Dabei wird kein Vergleich zwischen den Schulen gemacht, sondern versucht ein Muster aufzudecken, welches der Berufsvorbereitung generell zu Grunde liegt. Im Gegensatz dazu soll mit der Clusteranalyse versucht werden, Unterschiede zwischen den verschiedenen Schultypen festzustellen. Denn es ist durchaus eine interessante Frage, ob sich die verschiedenen Schultypen in unterschiedlichen Clustern widerspiegeln.

Bei der Durchführung der beiden Verfahren mit dem vorliegenden Datensatz werde ich viele der zuvor beschriebenen statistischen Analysemethoden anwenden. Alle Grafiken und Tabellen, die nicht direkt im Text dargestellt sind, können im Anhang unter A.3.b für die Faktorenanalyse und unter A.3.c für die Clusteranalyse gefunden werden.

---

<sup>27</sup> Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: S. 439

### 3.2.2.1. Faktorenanalyse

Ausgangspunkt der Faktorenanalyse ist die standardisierte Korrelationsmatrix<sup>28</sup> der zu untersuchenden Variablen. Sie dient dazu, die Eignung der Variablen für eine Faktorenanalyse zu ermitteln.

Dazu werden zuerst auf die Korrelationsmatrix die statistischen Mittel KMO-Kriterium und MSA-Kriterium angewandt.

KMO-Test	
Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.	,777

Abbildung 3.7.: KMO-Kriterium

Das KMO-Kriterium liegt für den vorliegenden Anwendungsfall bei 0,777 (Abbildung 3.7.) und weist somit eine “ziemlich gute”<sup>29</sup> Eignung der Daten für eine Faktorenanalyse hin. Der KMO-Wert sollte allerdings vorsichtig interpretiert werden, da das Kriterium eigentlich metrisch-skalierte Daten erfordert. Zielführender ist die Anwendung des MSA-Kriteriums, welches auf der Hauptdiagonalen der Anti-Image-Korrelations-Matrix abzulesen ist (Anhang A.3.b). Von den 14 Variablen, die endgültig für die Faktorenanalyse genutzt werden, haben 6 einen MSA  $> 0,8$  und somit eine “verdienstvolle”<sup>29</sup> Eignung. Ebenfalls 6 Variablen besitzen einen MSA-Wert  $> 0,7$ , was einer “ziemlich guten” Eignung entspricht. Zwei Variablen sind mit einem MSA  $> 0,6$  “mittelmäßig” geeignet. Mit der Anti-Image-Matrix, lässt sich ein weiteres Kriterium zur Überprüfung der Variableneignung untersuchen. Nach Dziuban und Shirkey<sup>30</sup> sollen nur maximal 25% der Nicht-diagonal-Elemente ungleich 0 ( $> 0,09$ ) sein. Da auch dieses Kriterium für den vorliegenden Datensatz deutlich erfüllt wird, scheinen die erhobenen Daten für eine Faktorenanalyse geeignet zu sein.

Jedoch ist an dieser Stelle anzumerken, dass von den für die Faktorenanalyse anfänglich vorgesehen 22 Variablen 8 aufgrund der zu geringen Eignung ausgeschlossen werden müssen. Im Vorfeld der Analyse ist zu berücksichtigen, dass das Vorliegen ordinaler und binärer Daten u. U. zu Problemen führen kann. Eventuell kann die

<sup>28</sup> Siehe Anhang A.3.b

<sup>29</sup> Abstufung der Eignungen: Vgl. Kaiser/Rice (1974): S. 111ff.

<sup>30</sup> Vgl. Dziuban/Shirkey (1974): S. 359

Skala von 1 bis 4 bzw. von 0 und 1 nicht breit genug gefächert sein, um plausible Ergebnisse zu erhalten.

Da die grundlegende Eignung der erhobenen Daten gegeben ist, werden als nächstes die Kommunalitäten bestimmt. Wie zuvor erklärt, kommen dazu die PCA und PAF zur Anwendung, sodass man die folgende Ergebnisse gemäß Abbildung 3.8. erhält:

	PCA		PAF	
	Kommunalität (Startwert)	Kommunalität (Endwert)	Kommunalität (Startwert)	Kommunalität (Endwert)
<b>oppind</b>	1,000	0,432	0,237	0,255
<b>thoughtgrad</b>	1,000	0,686	0,481	0,616
<b>job</b>	1,000	0,697	0,452	0,623
<b>prepschool</b>	1,000	0,647	0,423	0,539
<b>prepclass</b>	1,000	0,596	0,361	0,452
<b>schoolusef</b>	1,000	0,655	0,449	0,546
<b>schoolinfo</b>	1,000	0,586	0,343	0,425
<b>prepprivat</b>	1,000	0,339	0,248	0,228
<b>ambitious</b>	1,000	0,587	0,423	0,498
<b>classbest</b>	1,000	0,441	0,235	0,262
<b>diligent</b>	1,000	0,547	0,438	0,452
<b>organized</b>	1,000	0,497	0,435	0,393
<b>goaloriented</b>	1,000	0,601	0,512	0,536
<b>dedicated</b>	1,000	0,627	0,490	0,571

Abbildung 3.8.: Kommunalitäten-Matrix der angewandten Extraktionsverfahren

Wie bereits im theoretischen Teil beschrieben, wird bei der PCA der Anfangswert auf 1 gesetzt und bei der PAF das multiple Bestimmtheitsmaß als Ausgangswert für die Analyse angenommen. Allerdings ist zu erkennen, dass vor allem bei der PAF geringe Kommunalitäten-Werte vorliegen und bei der PCA große Abweichungen von den Startwerten entstanden sind. Daraus muss geschlossen werden, dass das weitere Durchführen der Faktorenanalyse mit diesen Variablen zumindest fragwürdig ist, denn bei einigen Variablen wird nur ein sehr geringer Teil der Varianz durch die Faktoren erklärt. Außerdem ist auffällig, dass einige Variablen, als Beispiel sei hier *prepprivat* genannt, zwar von dem KMO-Kriterium als "verdienstvoll" eingeschätzt



werden, aber den Kommunalitäten zufolge eher weniger für die Analyse geeignet sind.

Einhergehend mit der Berechnung der Kommunalitäten geht die Frage nach der Anzahl der Faktoren. Bekannt ist bereits, dass die maximale Faktorenanzahl genau der Anzahl der Variablen entspricht, damit jedoch der Sinn und Zweck einer Faktorenanalyse verfehlt sein würde. Der folgenden Tabelle 3.9. sind die extrahierten Eigenwerte der PCA und PAF mit und ohne Varimax-Rotation zu entnehmen. Nach dem Kaiser-Kriterium, das empfiehlt, Faktoren mit Eigenwert größer 1 zu extrahieren, wurden die folgenden drei Eigenwerte extrahiert.

PCA					PAF			
Komponente	Summe der quadr. Faktorladungen		Rotierte Summe d. quadr. Faktorl.		Summe der quadr. Faktorladungen		Rotierte Summe d. quadr. Faktorl.	
	Gesamt	kum. %	Gesamt	kum. %	Gesamt	kum. %	Gesamt	kum. %
1	3,682	26,297	3,117	22,263	3,162	22,589	2,565	18,324
2	2,449	43,793	2,431	39,624	1,944	36,477	1,926	32,079
3	1,805	56,686	2,389	56,686	1,288	45,677	1,904	45,677

Abbildung 3.9.: Eigenwerte der Extraktionsverfahren mit und ohne Rotation

Durch beide Verfahren mit Extraktion von Faktoren mit Eigenwert größer als 1 lässt sich nur rund die Hälfte der auftretenden Gesamtvarianz erklären. Alle vier Analysen extrahieren drei Faktoren, die auch bei allen Verfahren ungefähr gleich groß sind. Ein anderes Verfahren, mit dessen Hilfe die Anzahl der Faktoren festgelegt werden kann, ist der Screen-Test, der exemplarisch für die Hauptkomponentenanalyse mit Varimax-Rotation angewendet wird und in Abbildung 3.10. dargestellt ist.

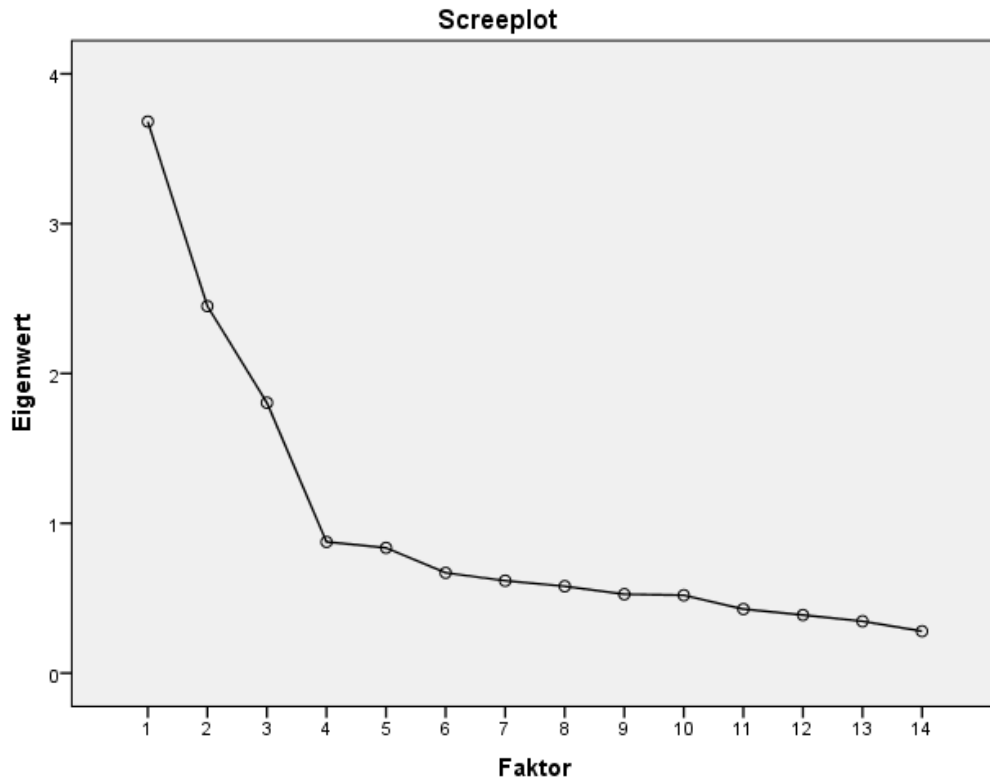


Abbildung 3.10.: Screen-Test

Auch wenn der größte Sprung zwischen den Eigenwerten der Komponenten 1 und 2 vorliegt und per Definition die Anzahl der Faktoren links des größten Gefälles gewählt werden sollte, könnte man auch rein optisch die Entscheidung treffen, analog zum Kaiser-Kriterium, 3 Komponenten zu extrahieren.

Nach der Festlegung der Anzahl der Faktoren müssen die Faktoren in Abhängigkeit der Variablen interpretiert werden. Welches “dahinterliegende” Merkmal die Faktoren darstellen, kann durch die Variablen festgestellt werden, auf die sie hoch laden. Zusammenhänge können mit Hilfe einer Faktorladungsmatrix festgestellt werden. Die folgende Matrix in Abbildung 3.11. führt die Faktorladungen der untersuchten 14 Variablen für die PCA und PAF mit und ohne Varimax-Rotierung zusammen.

	PCA						PAF					
	ohne Rotation			Varimax-Rotation			ohne Rotation			Varimax-Rotation		
	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3
oppi nd	0,416	0,043	<b>-0,507</b>	0,069	-0,029	<b>0,653</b>	0,363	0,030	-0,350	0,100	-0,013	0,495
thou ghtgr ad	<b>0,598</b>	0,066	<b>-0,569</b>	0,183	-0,008	<b>0,808</b>	<b>0,583</b>	0,058	-0,523	0,181	-0,004	<b>0,764</b>
job	0,458	0,181	<b>-0,674</b>	-0,007	0,077	<b>0,831</b>	0,450	0,168	<b>-0,626</b>	-0,002	0,075	<b>0,786</b>
prep priva t	<b>0,514</b>	0,082	-0,261	0,275	0,062	<b>0,509</b>	0,441	0,058	-0,173	0,257	0,053	0,399
prep scho ol	0,107	<b>0,796</b>	0,051	-0,009	<b>0,796</b>	0,118	0,104	<b>0,725</b>	0,045	-0,007	<b>0,726</b>	0,109
prep class	0,049	<b>0,769</b>	-0,042	-0,102	<b>0,749</b>	0,159	0,048	<b>0,670</b>	-0,037	-0,089	<b>0,652</b>	0,137
scho oluse f	0,046	<b>0,782</b>	0,203	0,024	<b>0,808</b>	-0,043	0,047	<b>0,715</b>	0,180	0,021	<b>0,738</b>	-0,034
scho olinf o	0,031	<b>0,704</b>	0,300	0,075	<b>0,749</b>	-0,141	0,028	<b>0,604</b>	0,243	0,059	<b>0,640</b>	-0,109
ambi tious	<b>0,668</b>	-0,129	0,352	<b>0,763</b>	-0,017	0,063	<b>0,619</b>	-0,119	0,316	<b>0,701</b>	-0,016	0,081
class best	0,386	-0,139	<b>0,522</b>	<b>0,622</b>	-0,012	-0,234	0,330	-0,112	0,374	0,494	-0,017	-0,130
dillig ent	<b>0,695</b>	-0,013	0,252	<b>0,714</b>	0,079	0,175	<b>0,636</b>	-0,014	0,217	<b>0,643</b>	0,070	0,184
orga nized	<b>0,639</b>	-0,154	0,254	<b>0,691</b>	-0,062	0,125	<b>0,575</b>	-0,133	0,212	<b>0,609</b>	-0,052	0,139
goal orien ted	<b>0,766</b>	-0,105	0,060	<b>0,685</b>	-0,043	0,361	<b>0,722</b>	-0,099	0,067	<b>0,645</b>	-0,037	0,344
dedic ated	<b>0,765</b>	-0,047	0,198	<b>0,749</b>	0,040	0,0253	<b>0,727</b>	-0,050	0,199	<b>0,714</b>	0,037	0,245

Abbildung 3.11.: Faktorladungsmatrix

Von all den vier dargestellten Varianten liefert die PCA mit Varimax-Rotation die besten und auch plausibelsten Ergebnisse. Dennoch ist festzustellen, dass beide Verfahren mit und ohne Rotation zu relativ ähnlichen Ergebnissen führen. Außerdem sei an dieser Stelle erwähnt, dass dies auch auf Rotationen zutrifft, die Korrelationen zwischen den Faktoren annehmen, also oblique Rotationen, wie beispielsweise die

Promax-Rotation. Der Vorteil bei der PCA mit Varimax-Rotation ist, dass alle Variablen nur auf einen Faktor hoch geladen sind und dies für alle Variablen zutrifft, also es gibt keine Variable, die keinem Faktor zugeordnet werden kann und es gibt keine Variable, die mehreren Faktoren zugeordnet werden kann. Außerdem ist bei der PCA mit Varimax-Rotation die inhaltliche Interpretation sehr sinnvoll. Deshalb betrachte ich in den folgenden Ausführungen nur noch diese Variante.

Die Variablen *oppid*, *thoughtgrad*, *job* und *prepprivat* beziehen sich alle auf Berufsvorbereitungen, die jeder Einzelne selbst betreiben kann, indem er durch außerschulische Aktivitäten und Recherchen seine Interessen erkundet und sich über seine Berufswünsche Gedanken macht. Dieser Faktor -Faktor 3- kann unter dem Name "private Berufsvorbereitung" zusammengefasst werden. Die Variablen *prepschool*, *prepclass*, *schoolusef* und *schoolinfo* drücken die Informationsmöglichkeiten aus, die in der Schule angeboten werden. Sie alle werden unter Faktor 2 als "schulische Berufsvorbereitung" gebündelt. Faktor 1 beinhaltet die Variablen *ambitious*, *classbest*, *diligent*, *organized*, *goaloriented* und *dedicated*. Sie können als "Charakterzüge" oder "Motivation" interpretiert werden. Demnach hängt das Maß an Berufsvorbereitung auch von jedem Einzelnen ab, beispielsweise, inwiefern er motiviert ist, sich mit dem Thema auseinander zu setzen.

Zur Veranschaulichung der drei Faktoren, werden die Variablen mit den jeweiligen Faktorladungen in dem Diagramm Abbildung 3.12. darstellt.

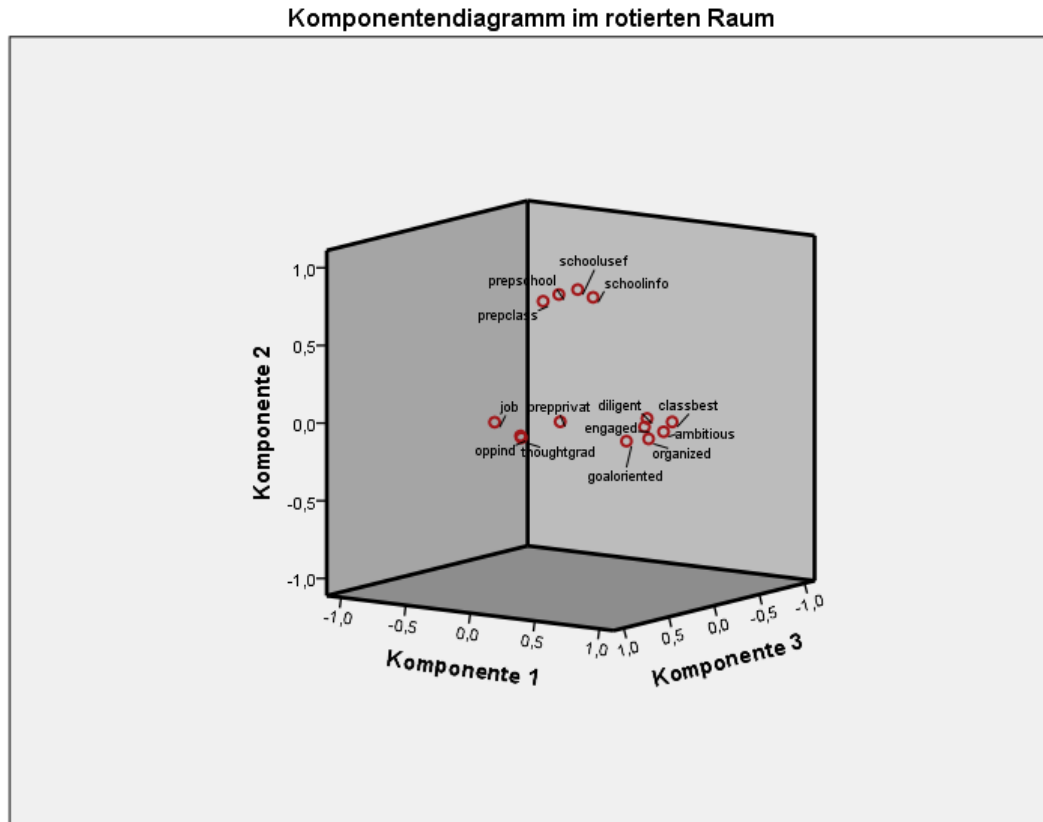


Abbildung 3.12.: Komponentendiagramm im rotierten Raum

Die Abbildung zeigt die Variablen mit den Faktorladungen der Varimax-Rotation einer PCA. Man erkennt deutlich die Clusterung der Variablen zu den 3 dahinter liegenden Komponenten.

Für die folgenden Analysen werden die jeweiligen zu den Faktoren gehörenden Variablen zu einer Variablen mit Summenskala zusammengefasst und durch diesen neuen Faktor im Datensatz ersetzt. Die neu in den Datensatz aufgenommenen Variablen heißen *Berufsvorbereitung Schule* (0,789), *Berufsvorbereitung Privat* (0,703) und *Leistungsorientierte Persönlichkeit* (0,809). Zusätzlich dazu wird für jede Gruppe von Variablen Cronbachs Alpha<sup>31</sup> bestimmt. Die daraus resultierenden Werte stehen in Klammern hinter den Faktoren bzw. neuen Variablen. Für die so zusammengefassten Variablen ergibt sich zweimal die Bewertung “akzeptabel” und einmal die Bewertung “gut”<sup>32</sup>.

<sup>31</sup> Cronbachs Alpha ist ein Maß für die Eignung der Faktoren.

<sup>32</sup> Darren George, Paul Mallery: SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update. 4. Auflage. Allyn & Bacon (2002)

### 3.2.2.2. Clusteranalyse

Im Anschluss an die Faktorenermittlung soll versucht werden, homogene Gruppen innerhalb des Datensatzes nachzuweisen. Dazu sollen die beiden agglomerativen hierarchischen Verfahren Single-Linkage und Ward angewendet werden.

Zu Beginn einer Clusteranalyse wird eine Distanz- oder Ähnlichkeitsmatrix über alle Fälle aufgestellt. Die Distanzmatrix<sup>33</sup> wird mit Hilfe der quadrierten Euklidischen Distanz berechnet. Normalerweise sollten für die Euklidische Distanz metrisch-skalierte Variablen vorliegen. Dennoch wurde die Euklidische Distanz zur Berechnung der Distanzen für die vorliegenden ordinalen bzw. binären Variablen verwendet, da sie für das angewendete Ward-Verfahren besonders empfohlen wird. Allerdings kann es in diesem Fall möglich sein, u. U. nicht aussagekräftige Ergebnisse zu erhalten.

In die Distanzmatrix und in die Analyse werden anders als bei der Faktorenanalyse alle Variablen aufgenommen; außer demographische Informationen wie Alter, Geschlecht und Schultyp, so kann im Nachhinein bei der Clusterfestlegung bestimmt werden, inwiefern diese beschreibenden Kriterien "sich selbst sortieren".

Als nächstes wird eine Single-Linkage Clusteranalyse durchgeführt mit der oben erwähnten quadrierten Euklidischen Distanz zur Distanzermittlung. So können vorab mögliche Ausreißer aus der Analyse genommen werden. Der dafür entscheidende Teil des Dendrogramms ist in Abbildung 3.13. dargestellt.

---

<sup>33</sup> Aufgrund der hohen Anzahl an Fällen, kann die Distanzmatrix (228x228) in dieser Arbeit nicht so dargestellt werden, dass man sie sinnvoll nachvollziehen könnte. Falls nötig, muss sie mit Hilfe des Datensatzes bestimmt werden.

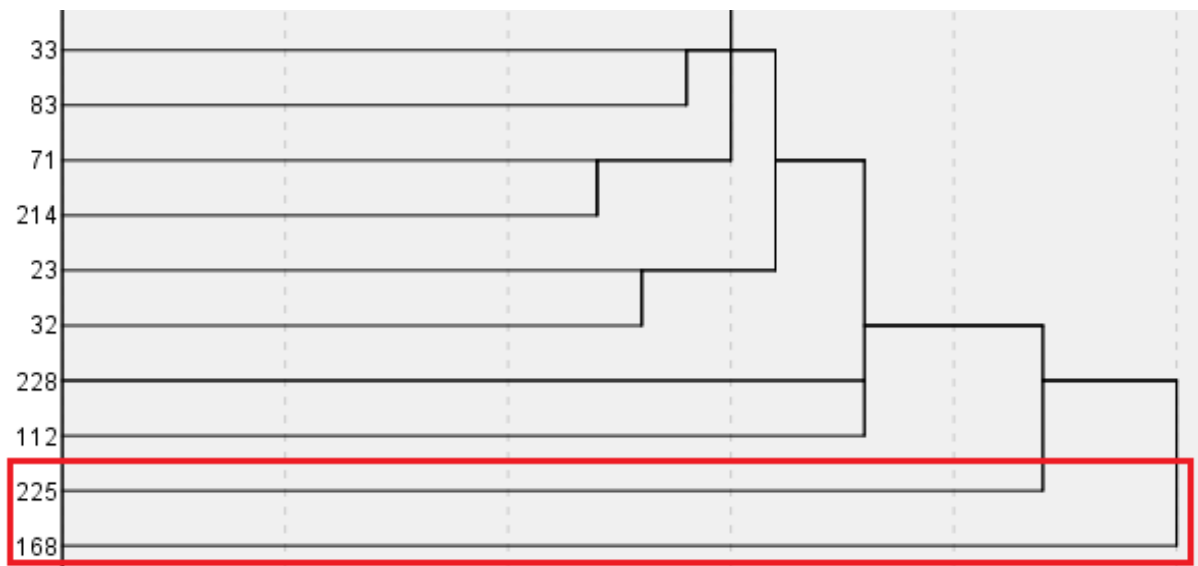


Abbildung 3.13.: Dendrogramm Single-Linkage

Die Grafik zeigt die letzten Clusterzusammenlegungen. Man kann deutlich in dem roten Kasten erkennen, dass die Fälle 225 und 168 nur bei den letzten Analyseschritten mit einem anderen Cluster zusammengelegt werden und somit als Ausreißer betrachtet werden können. Diese Fälle wurden deshalb von den weiteren Untersuchungen ausgeschlossen.

Anschließend wird eine Clusteranalyse mit dem Ward-Verfahren durchgeführt.

#### Zusammenfassung der Fallverarbeitung

Fälle					
Gültig		Fehlend		Gesamt	
N	Prozent	N	Prozent	N	Prozent
203	89,0	25	11,0	228	100,0

a. Ward-Linkage

Abbildung 3.14.: Zusammenfassung Ward-Verfahren

Wie der Abbildung 3.14. zu entnehmen ist, gibt es insgesamt 228 zu untersuchende Fälle, da 2 aus der Analyse herausgenommen wurden. 25 der 228 Fälle werden ebenfalls automatisch von der Analyse ausgeschlossen, weil bei der Datenerhebung mindestens eine Variable nicht ausgefüllt wurde. Gemäß der vorangegangenen Betrachtungen werden die Distanzen mit Hilfe der quadrierten Euklidischen Distanz berechnet. Im Anhang unter A.3.c findet man die komplette Zuordnungsübersicht mit jedem einzelnen Fusionierungsschritt. Das dazugehörige Dendrogramm ist in Abbildung 3.15. dargestellt.

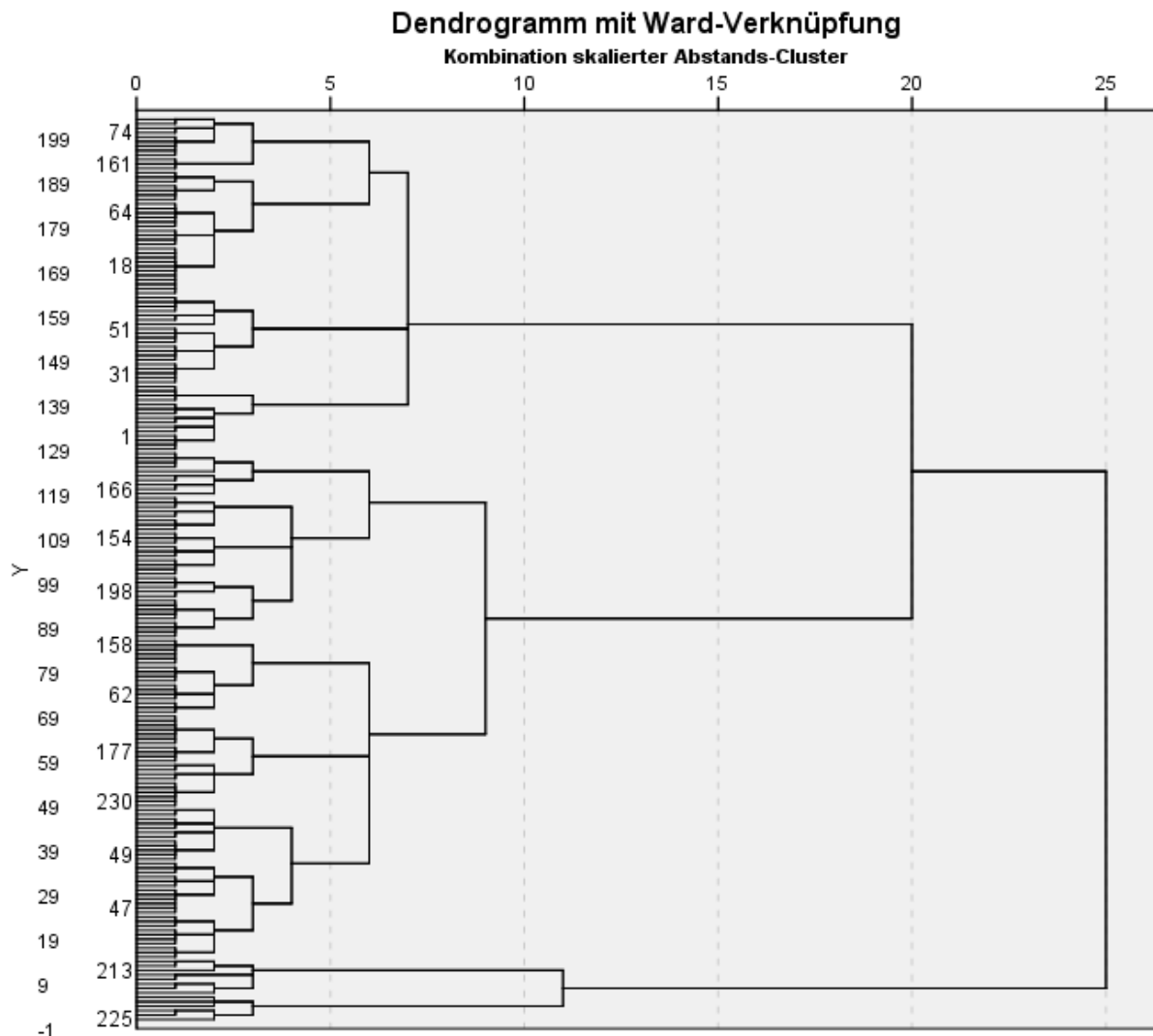


Abbildung 3.15.: Dendrogramm Ward-Verfahren

Anhand der Zuordnungsübersicht muss nun versucht werden, die Anzahl der Cluster festzulegen. Aufgrund der Vielzahl von Fällen kann das Elbow-Kriterium nicht plausibel angewendet werden. In der Zuordnungsübersicht ist erkennbar, dass bei dem letzten Fusionierungsschritt beim Zusammenlegen von Cluster 1 und 23 der größte Sprung im Koeffizienten (bei dem Ward-Verfahren die Fehlerquadratsumme) vorliegt. Dies könnte als Argument dafür gelten, die untersuchten Objekte in zwei Cluster einteilen. Dennoch habe ich mich bei zusätzlicher Betrachtung des Dendrogramms in Abbildung 3.15. entschieden, drei Cluster für die weiteren Analysen zu verwenden. Die ausgewählten Cluster sind in Abbildung 3.16. markiert.



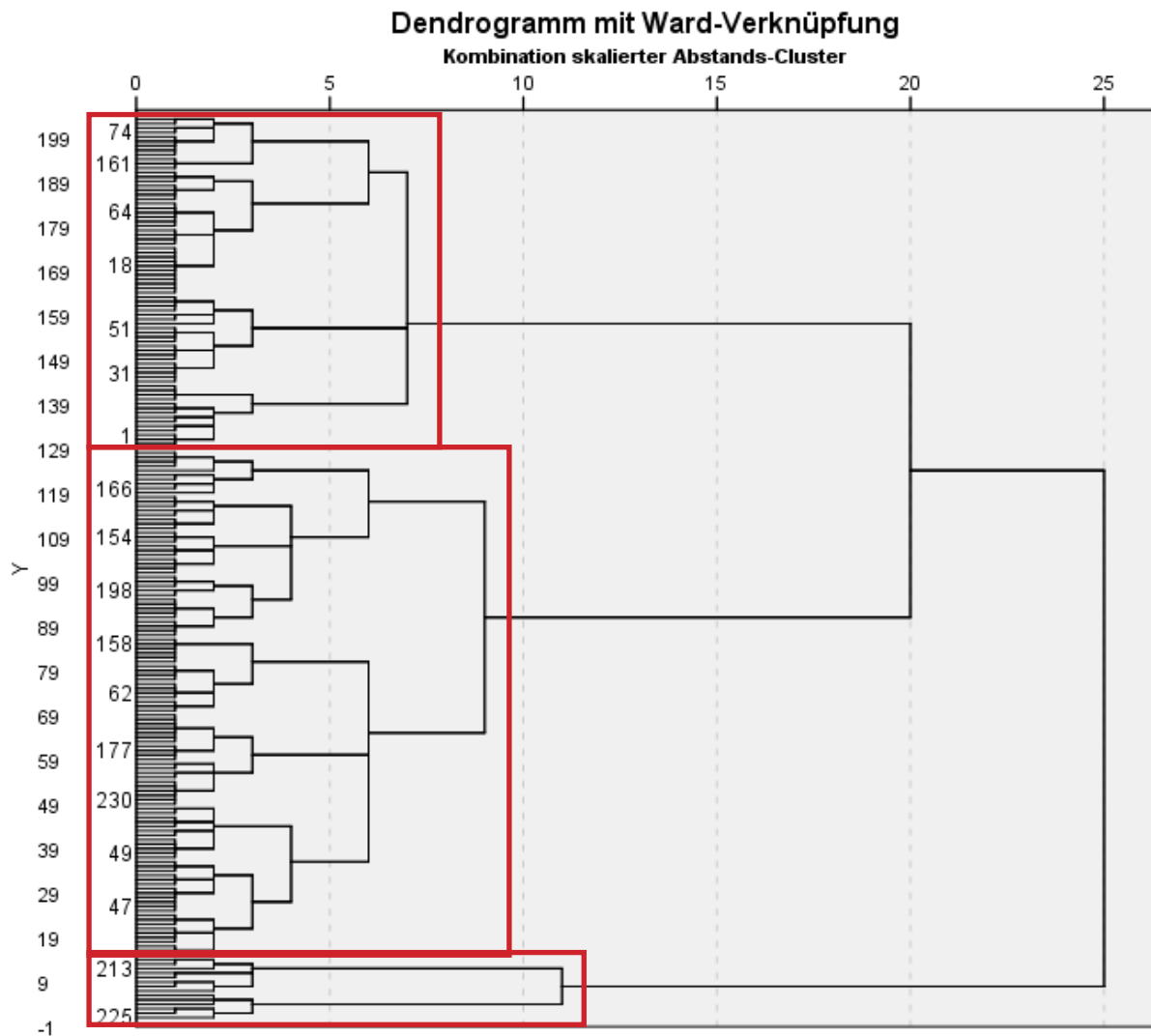


Abbildung 3.16.: Dendrogramm Ward-Verfahren mit markierten Clustern

Diese Entscheidung ist konform mit dem Test von Mojena (Abbildung 3.17.). Rechnet man mit den Koeffizientenwerten der Zuordnungsübersicht, so wären Lösungen mit 2-4 Clustern möglich, da bei allen der standardisierte Fusionkoeffizient über dem eher konservativen Wert 2,5 liegt. Wählt man den, wie unter 3.2.1.2. dargestellt, von Mojena selbst vorgeschlagenen Grenzwert von 2,75, erhält man eine Lösung mit 3 Clustern.

Fusionsschritt	1	...	199	200	201
Clusterzahl	202	...	4	3	2
Fusionskoeffizient a	3.806.333	...	3.109.275	3.242.345	3.488.813
Standardisierte Fusionskoeffizienten $A_i$	-1,20	...	2,59	2,75	3,06

Abbildung 3.17.: Test von Mojena

Zur Klarstellung soll noch angemerkt werden, dass Cluster 1 das oberste, Cluster 2 das mittlere und Cluster 3 das unterste der drei Cluster ist. Im Rahmen der Analyse werden für die frei erhaltenen Cluster zunächst je Gruppe die Durchschnitte über alle Variablen gebildet, so dass für den Vergleich der Cluster eine Basis geschaffen wird. Anschließend wird je Variable die Effektstärke mit Hilfe des Cohens  $f^2$  berechnet. Dadurch lassen sich Aussagen darüber treffen, ob Unterschiede zwischen den Clustern bestehen und wenn ja, wie groß diese sind bzw. durch welche Variablen ein Cluster charakterisiert wird. Die Tabelle in Abbildung 3.18. zeigt die Anzahl der Personen, die sich jeweils im Cluster befinden, die jeweiligen Durchschnitte der Variablen und in der vierten Spalte ist die Effektstärke aufgeführt. Cohen hat kritische Werte festgelegt, an die man sich bei der Interpretation der Effektstärke<sup>34</sup> halten kann. Demnach spricht man bei Werten größer als oder gleich 0,4 von starken Effekten, in dieser Tabelle mit dunkelroter Füllung kenntlich gemacht. Mittlere Effekte (Werte größer oder gleich 0,25) sind an dem mittleren Rotton erkennbar. Bei Werten größer oder gleich 0,10 spricht man von einem leichten Effekt. Diese sind mit einer hellroten Füllung markiert. Werte kleiner als 0,10 können nicht als effektvoll angenommen werden. Die Tabelle in Abbildung 3.18. enthält eine Auswahl an Variablen, die mindestens einen leichten Effekt aufweisen und als sinnvoll für die Charakterisierung der Cluster angesehen werden können. Die rot markierten Zahlen in den Clusterspalten sollen die Variablen und die Clusterdurchschnitte hervorheben, die für die Interpretation interessant sind. Weshalb für die Untersuchung der Cluster hinsichtlich etwaiger Unterschiede die Effektstärke betrachtet wird, kann im Abschnitt 3.4. nachgelesen werden.

Variablen	Cluster 1	Cluster 2	Cluster 3	Effektstärke*
Personen in dem Cluster	75	114	13	
school_type_b	75%	24%	36%	0,44
age	15,39	15,88	15,64	0,25
sex_b	61%	46%	62%	0,16
avrggrade_t	2,49	2,59	2,64	0,11
degree_abi	91%	60%	57%	0,34
degree_msa	0%	11%	14%	0,24
degree_msamehr	8%	26%	14%	0,21
degree_eBB	1%	3%	14%	0,39

Abbildung 3.18.: Effektstärken für Cluster - Fortsetzung auf der nächsten Seite

<sup>34</sup> Interpretation der Effektstärke nach: Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2. Aufl., Hillsdale: Lawrence Erlbaum Associates (1998)

Variablen	Cluster 1	Cluster 2	Cluster 3	Effektstärke*
Biologie	25%	23%	14%	0,13
Mathe	21%	8%	14%	0,17
Physik	13%	4%	7%	0,14
Informatik	13%	4%	7%	0,16
aftergrad_uni	88%	55%	50%	0,36
aftergrad_ausbildung	8%	30%	50%	0,43
Elektronikindustrie	8%	11%	21%	0,20
Chemieindustrie	8%	11%	29%	0,33
BankenundVersicherungen	32%	15%	21%	0,17
InformationsKommunikationstechnologien	12%	8%	29%	0,33
prepmore	2,04	2,34	2,64	0,33
engfriends	2,13	1,68	1,50	0,35
numberpriv_t	0,64	1,09	5,21	1,50
numberscl_t	0,27	1,70	3,00	0,91
Politik	12%	5%	14%	0,16
WirtschaftINT	19%	11%	7%	0,15
BildungINT	8%	5%	0%	0,15
preptotal	3,64	2,85	2,29	0,55
Faktor1BerufsvorbereitungSchule	3,32	2,70	3,08	0,40
Faktor2BerufsvorbereitungPrivat	2,28	1,90	1,52	0,47
Faktor3LeistungsorientiertePersönlichkeit	2,19	2,00	1,76	0,32
* zur Berechnung der Effektstärke wurde Cohens $f^2$ verwendet				

Abbildung 3.18.: Effektstärken für Cluster

Zum Verständnis der Prozentwerte muss gesagt werden, dass solche Variablen in Dummy-Variablen umgewandelt werden und der Prozentsatz den Anteil der Befragten angibt, bei denen das untersuchte Kriterium zutrifft. Ich möchte nochmals erwähnen, dass die Variablen *schooltype* und *sex* nicht in der Clusternalyse berücksichtigt werden und trotzdem lässt sich zumindest bei der Variable *schooltype* eine starke Trennung feststellen.

Cluster 1 besteht zu 75% aus Gymnasiasten und damit einhergehend gibt es einen hohen Anteil an Schülern, die Abitur machen und anschließend eine Universität besuchen wollen. Da die Anteile an Gymnasiasten und Schülern, die Abitur machen wollen, nicht die gleichen Werte aufweisen, kann vermutet werden, dass diesem Cluster auch einige Schüler angehören, die eine ISS besuchen und dort ihr Abitur machen wollen. Wie durch die farbliche Unterlegung in der rechten Spalte (Effekt-

stärke) sichtbar wird, liegt bei dem Schultyp ein starker Effekt, bei dem angestrebten Schulabschluss und dem Universitätsbesuch ein mittlerer Effekt vor. Daraus kann man schließen, dass bei den drei Variablen aussagekräftige Unterschiede vorliegen.

Bei den Fächern *Mathematik*, *Physik*, *Biologie* und *Informatik* ist ein durchgehend leichter Effekt erkennbar. Er darauf hindeutet, dass im Cluster 1 ein höherer Prozentsatz an naturwissenschaftlich interessierten Schülern zu finden ist. Ein leichter Effekt lässt sich außerdem bei der Branchenwahl erkennen. Besonders hoch ist der Anteil an Schülern, die sich vorstellen können, in der Banken- und Versicherungsbranche zu arbeiten. Dies kann möglicherweise dadurch erklärt werden, dass in diesen Bereich das Abitur oder sogar ein Universitätsabschluss von Nöten ist. Zu den Interessen der Schüler aus Cluster 1 lässt sich ferner feststellen, dass es so scheint, als seien diese Schüler im Vergleich zu denen der anderen Clustern besonders an den Themen Politik und Wirtschaft interessiert. Dem liegen zwar nur kleine Effekte zu Grunde, aber zumindest für das Interessensgebiet Wirtschaft lässt sich dies mit der Branchenwahl verbinden.

Nicht besonders günstig sieht es dagegen bei den Schülern des Clusters 1 aus, wenn es um die Vorbereitung auf das Leben nach dem Schulabschluss geht. Nach den Angaben der Schüler zu urteilen, sind sie von den drei Clustern insgesamt am schlechtesten auf das Berufsleben vorbereitet, wobei sich bei den entsprechenden Variablen viele starke Effekte finden lassen. Beispielsweise haben diese Schüler von allen am wenigsten Veranstaltungen zur Berufsvorbereitung besucht, egal ob sie schulischer oder privater Natur sind. Sie schätzen auch die Berufsvorbereitung ihrer Schule - abgesehen durch den Faktor 1 - am schlechtesten ein. Dem gegenüber steht mit einem mittleren Effekt die Erkenntnis, dass eventuell eine bessere Vorbereitung nötig wäre (*prepmore*). Allerdings sollte an dieser Stelle bedacht werden, und dieser Gedankengang wird in der Zusammenfassung nochmals aufgegriffen, dass es für Schüler des Gymnasium in der 10. Klasse noch nicht besonders dringlich ist, sich intensiv auf das Berufs- bzw. Universitätsleben vorzubereiten.

Interpretationen zu den Ergebnissen der Cluster 2 und 3 sind deutlich schwieriger. Das liegt vor allem daran, dass Cluster 2 bei vielen Variablen weder den Minimum- noch den Maximumwert darstellt und sich in Cluster 3 nur dreizehn Personen wiederfinden. Das einmalige Auftreten einer Variablenausprägung kann dort eventuell schon

große Unterschiede zu den anderen Gruppen bewirken. Dennoch kann festgestellt werden, dass sich im Cluster 2 hauptsächlich Schüler der ISS befinden. Gleichzeitig lässt sich ein gesteigerter Anteil (37%) an Schülern erkennen, die zunächst einmal nach dem MSA ihre Schulausbildung beenden möchten. Trotzdem ist auch in diesem Cluster der vorrangig gewünschte Schulabschluss das Abitur. Das Phänomen, dass in allen Clustern das Abitur den höchsten Anteil ausmacht, lässt sich damit begründen, dass der an einem Gymnasium erhaltene Schulabschluss üblicherweise das Abitur ist, es an der ISS allerdings auch abgelegt werden kann. Es ist also nicht erstaunlich, dass von allen Befragten mehr als die Hälfte das Abitur als Schulabschluss anstrebt. Eine Besonderheit des Cluster 2 ist, dass sich dort mehr Mädchen als Jungen befinden.

Gegenüber Cluster 1 lässt sich zudem eine bessere Berufsvorbereitung erkennen. Zum Einen besuchen die Schüler mehr Veranstaltungen und zum Anderen werden die privaten (*Faktor 2*) und schulischen (*Faktor 1*) Maßnahmen als besser eingestuft. Dies führt insgesamt zu einem besseren Befinden bezüglich der Gesamtvorbereitung (*preptotal*).

Wie bereits erwähnt, ist Cluster 3 eine besonders kleine Gruppe, überwiegend bestehend aus Schülern der ISS. Der Anteil der Schüler, die Abitur oder MSA machen wollen, ist im Vergleich zu Cluster 2 deutlich niedriger. Dagegen gibt es einen gesteigerten Prozentsatz (14%) an Schülern, die als Bildungsabschluss den eBB erlangen möchten. Aufgrund der geringen Anzahl an Personen ist fraglich, ob diese Information für die Interpretation des Clusters von Bedeutung ist. Außerdem lässt sich ein Muster in den von diesen Schülern gewählten Branchen erkennen. Mit den Branchen Elektronik (21%), Chemie (29%) und Informations- und Kommunikationstechnik (29%) kann zumindest ein Interesse an Technik vermutet werden.

Ebenfalls hoch ist in Cluster 3 der Anteil an Schülern, die nach dem Schulabschluss eine Ausbildung beginnen möchten. Auf jeden Fall charakterisierend für dieses Cluster ist eine intensive Berufsvorbereitung. Sowohl für die Anzahl an privaten als auch an schulischen Veranstaltungen zur Berufsvorbereitung liegen starke Effekte vor. Dies trifft auch für die Bewertung der Vorbereitung auf das Berufsleben im Großen und Ganzen zu. Die Schüler des Clusters 3 geben für diese Frage eine Durchschnittsnote von 2,29 (auf einer Schulnotenskala) an. Gemäß der Schülerangaben

befinden sich im Cluster 3 die leistungsorientiertesten Persönlichkeiten (*Faktor 3*). Dabei ist fraglich, inwiefern dieses Ergebnis plausibel ist, denn es liegt zumindest ein logischer Widerspruch vor, wenn Schüler, die größtenteils ihre Schullaufbahn mit dem MSA beenden, sich am engagiertesten, fleißigsten und zielorientiertesten einschätzen. Eventuell ist diese (mögliche) Verzerrung auf die kleine Clustergröße zurückzuführen.

Zusammenfassend lässt sich sagen, dass Cluster 1 hauptsächlich aus Gymnasiasten besteht, der Großteil der Schüler Abitur machen und eine Universität besuchen möchte. Zudem ist für sie eine relativ schlechte Vorbereitung auf das Berufsleben charakterisierend. Cluster 2 besteht größtenteils aus Schülern der ISS. Sie sind besser, insbesondere durch die Schule auf das Berufsleben vorbereitet. Cluster 3 zeichnet sich durch die besonders leistungsorientierten und gut auf das Berufsleben vorbereiteten Schülern aus. Hier sind Schultypen ungefähr gleich stark vertreten.

Führt man die Analysen nochmals durch, fasst jedoch nicht die aus der Faktorenanalyse bekannten Komponenten zusammen, erhält man nicht zwangsläufig die gleichen Ergebnisse. So sind zum Beispiel die Anzahl der Personen je Cluster ohne Zusammenfügen der Variablen ungefähr gleich groß. Insgesamt lassen sich die Ergebnisse ohne Anwendung der Faktoren besser interpretieren. Daher kann man zu dem Schluss kommen, dass die Faktorisierung der 14 Variablen gemäß Faktorenanalyse mit einem Informationsverlust verbunden ist, der zumindest zu leichten Veränderungen in den Ergebnissen führt.

### 3.3. Weitere Analysen

In diesem Abschnitt wird hauptsächlich untersucht, ob und inwiefern sich die Berufsvorbereitungsmaßnahmen und auch die Schüler der beiden Schultypen unterscheiden. Dazu werden für die beiden Gruppen wie bei der Clusteranalyse die Mittelwerte über alle Variablen je Schultyp mittels der Effektstärke untersucht. Da es sich um zwei Gruppen handelt, wird die Effektstärke vorzugsweise nach Cohens  $d^{35}$  ermittelt.

---

<sup>35</sup> Interpretation der Effektstärke nach: Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2. Aufl., Hillsdale: Lawrence Erlbaum Associates (1998)

Die Ergebnisse lassen sich in der Tabelle (Abbildung 3.19.) ablesen. Die Tabelle enthält solche Variablen, für die ein Effekt festgestellt werden kann und deren Interpretationen plausibel sind.

Variable	Gymnasium	ISS	Effektstärke*
Personen in der Gruppe	97	133	
age	15,24	16,02	1,11
sex	66%	40%	0,54
avrgrade_t	2,37	2,68	0,50
degree_abi	93%	51%	1,04
degree_msa	0%	13%	0,54
degree_msamehr	7%	29%	0,59
degree_eBB	0%	5%	0,33
Biologie	29%	19%	0,24
Mathe	19%	10%	0,25
Informatik	15%	2%	0,51
aftergrad_uni	89%	48%	0,96
aftergrad_ausbildung	9%	35%	0,66
aftergrad_sonstiges	1%	6%	0,27
Pflegebereich	4%	14%	0,36
BankenundVersicherungen	30%	12%	0,45
Eventmanagement	19%	31%	0,29
numberpriv_t	0,98	1,54	0,33
numberscl_t	0,53	1,91	1,14
Politik	16%	3%	0,46
StarsberühmtePersonen	5%	14%	0,31
Videospiele	35%	26%	0,21
Shoppen	8%	24%	0,44
WirtschaftINT	23%	5%	0,52
preptotal	3,28	3,05	0,22
Faktor1BerufsvorbereitungSchule	3,33	2,64	1,17
Faktor2BerufsvorbereitungPrivat	2,18	1,90	0,41
Faktor3LeistungsorientiertePersönlichkeit	2,08	2,04	0,07
* zur Berechnung der Effektstärke wurde Cohens d verwendet			

Abbildung: 3.19.: Unterschiede nach Schultypen

Die Markierungen in der Spalte “Effektstärke” haben dieselbe Bedeutung wie in der analog verwendeten Tabelle in Abbildung 3.18. Eine dunkelrote Füllung steht für einen starken Effekt, eine mittelrote für einen mittleren Effekt und die hellrote Füllung bedeutet, dass bei der untersuchten Variablen ein kleiner Effekt vorliegt. Alle Varia-

blen, deren Durchschnittswerte mittels einer Prozentzahl ausgedrückt werden, lassen sich als Anteil der Schüler je Schultyp deuten, auf die das jeweilige Kriterium zutrifft.

Wie der Tabelle zu entnehmen ist und bereits aus der Deskriptiven Statistik bekannt ist, haben an dem Gymnasium mehr Jungen und an der ISS mehr Mädchen an der Umfrage teilgenommen. Außerdem ist auffällig, dass ein starker Effekt bei dem Alter der befragten Schüler vorliegt, und die Gymnasiasten im Durchschnitt jünger sind. Dies lässt sich mit der "Schnellläufer-Schiene" des untersuchten Gymnasiums begründen, denn einige Schüler haben infolgedessen einen Jahrgang übersprungen.

Erwartungsgemäß möchte am Gymnasium der Großteil der Schüler das Abitur machen. Dementsprechend hoch ist auch der Anteil an Schülern die einen Universitätsabschluss erlangen wollen. Im Umkehrschluss ist an der ISS der Anteil der Schüler höher, die zunächst den MSA machen und anschließend eine Ausbildung beginnen wollen. Für diese Variablen liegen mittlere bis starke Effekte vor. Erstaunlich ist dennoch, dass auch mehr als die Hälfte der befragten 10. Klässler der ISS das Abitur anstreben wollen. Inwiefern sie die dafür benötigten Zulassungsvoraussetzungen erfüllen werden, ist im Vorfeld nicht absehbar.

Auch bei den naturwissenschaftlichen Schulfächern lässt sich - wie bei der Clusteranalyse - ein leichter bis mittlerer Effekt erkennen. Demnach ist an dem Gymnasium das naturwissenschaftliche Interesse der Schüler etwas größer. Doch wie bereits angedeutet, kann dies auch ganz einfach darauf zurückzuführen sein, dass das Geschlecht in den beiden Gruppen unbalanciert ist. Analog dazu lassen sich die unterschiedlichen Interessensgebiete erklären. Während ein leichter Effekt bei dem größeren Anteil der *Videospiel*-Interessierten an dem Gymnasium zu finden ist, liegt ein solcher an der ISS bei den Bereichen *Shoppen* und *Stars/berühmte Personen* vor. Es ist wohl anzunehmen, dass diese Charakterisierungen eher auf das vorherrschende Geschlecht je Gruppe als auf die Gesamtheit der Schüler je Schultyp zurückzuführen sind.

Weitere Unterschiede lassen sich in der Branchenwahl festmachen. Während sich die Gymnasiasten eher vorstellen könnten, in der Banken- und Versicherungsbranche zu arbeiten, tendieren die Schüler der ISS mehr zum Eventmanagement und Pflegebereich. Ähnliche Unterschiede liegen in der Clusteranalyse vor und lassen sich äquivalent interpretieren: Die Banken- und Versicherungsbranche ist mit dem



Abitur oder einem Universitätsabschluss zugänglicher und im Pflegebereich bzw. Eventmanagement hat man gute Perspektiven mit dem MSA. Da dieser Erkenntnis nur ein leichter Effekt zugrunde liegt, sollte man diese Schlussfolgerung mit Vorsicht betrachten.

Starke Effekt liegen bei den Schülern der ISS vor, wenn es um die Berufsvorbereitung geht. Die Schüler der ISS schätzen ihre privaten (Faktor 2), vor allem aber auch die schulischen Maßnahmen (Faktor 1) viel besser ein als Gymnasiasten. Dies erklärt das bessere Ergebnis bei der Gesamtbeurteilung (*preptotal*). Insgesamt kann angenommen werden, dass die Schüler der 10. Klassen an der ISS grundsätzlich besser vorbereitet sind. Ob zu diesem Zeitpunkt dies am Gymnasium überhaupt nötig ist, wird in der Zusammenfassung hinterfragt.

Zu Faktor 3 - dem leistungsorientierten Charakter - lässt sich sagen, dass er für die Schüler beider Schultypen im Mittel ungefähr gleich ausgeprägt ist. Er wird in die Tabelle aufgenommen, weil er gegenüber den anderen Variablen durch seine Extrahierung in der Faktorenanalyse eine gewisse Sonderstellung besitzt.

Alles in allem lassen sich Aussagen über die Schüler des jeweiligen Schultyps treffen, die vorher so erwartet werden konnten. Da die ISS einen höheren Anteil an Schulabgängern nach der 10. Klasse hat, sollten sie ihre Schüler besonders gut vorbereiten. Dass am Gymnasien mehr Schüler das Abitur ablegen und eine Universität besuchen wollen, ist ebenfalls keine Überraschung. Eventuell nicht vorherzusehen ist der ausgeglichene Wert für den leistungsorientierten Charakter, denn an der ISS befinden sich zu einem großen Teil Schüler mit Haupt- oder Realschulempfehlung (letzter Jahrgang vor der Strukturreform). Sie könnten gegenüber den Gymnasiasten möglicherweise weniger motiviert oder zielstrebig sein, was offensichtlich nicht zutreffend ist.

### 3.4. Repräsentativität der Umfrage

Im letzten Punkte der Datenanalyse wird hinterfragt, wie Aussagekräftig die durchgeführten Analysen und die daraus abgeleiteten Schlussfolgerungen sind. Wie schon bei der Clusteranalyse angedeutet, wurden die Effektstärken nicht ohne Grund ver-

wendet. Denn die Verwendung von Tests ist für die Ermittlung von signifikanten Unterschieden an dieser Stelle nicht geeignet. Dies liegt daran, dass Tests stets von der Stichprobe ausgehend Aussagen über die Grundgesamtheit treffen. Bei dieser Umfrage liegt keine Repräsentativität vor und es wäre somit nicht sinnvoll, Aussagen über die Grundgesamtheit treffen zu wollen. Die mangelnde Repräsentativität ist darauf zurückzuführen, dass bei der Auswahl der Teilnehmer der Umfrage keine Zufallsauswahl vorgelegen hat. Selbst wenn dieses Kriterium erfüllt wäre, müssten mehr als eine Schule je Schultyp befragt werden. Idealerweise müssten an mehreren zufällig ausgesuchten Schulen Stichproben an Schülern gezogen werden, die an der Umfrage teilnehmen.

Dennoch ist festzuhalten, dass die ermittelten Ergebnisse für die beiden vorliegenden Schulen durchaus valide sind und sinnvolle Schlussfolgerungen für die beiden Schulen möglich sind. Zu beachten ist allerdings, dass die Ergebnisse nicht für eine Verallgemeinerung herangezogen werden können.

## **4. Zusammenfassung**

Zum Schluss sollen nochmals die wichtigsten Aussagen zusammengefasst, aber auch eine Beurteilung der erhaltenen Ergebnisse abgegeben werden.

Durch die angewendeten Analyseverfahren konnten die die Berufsvorbereitung maßgeblich beeinflussenden Faktoren bestimmt werden. Die untersuchten Variablen konnten zu den drei Faktoren private Berufsvorbereitung, schulische Berufsvorbereitung und Charakterzüge zusammengefasst werden. Den Auswertungen folgend ist es zwar hilfreich, wenn Schüler in der Schule durch verschiedene Veranstaltungen auf das Berufsleben vorbereitet werden, jedoch ist der Einzelne für den Großteil der Vorbereitung selbst verantwortlich.

Des Weiteren konnten die Schüler schulübergreifend in drei relativ homogene Gruppen eingeteilt werden. Zwei Cluster kennzeichneten sich vor allem durch einen hohen Anteil an Schülern eines bestimmten Schultyps aus. Das dritte Cluster bestand aus Schülern, die vor allem durch ihre exzellente Berufsvorbereitung und ihre hohe Motivation bzw. ihren leistungsorientierten Charakter auffielen. Betont werden muss

allerdings, dass sich die Ergebnisse mit Anwendung der Faktoren von den Ergebnissen ohne Anwendung der Faktoren teilweise unterscheiden. Besonders in Cluster 3 wird die Veränderung deutlich sichtbar und die Auswertung des Clusters schwieriger.

Unter Punkt 3.3. wurden bestehende Unterschiede zwischen dem Gymnasium und der ISS festgestellt. Die erhaltenen Ergebnisse decken sich mit im Voraus zu erwartenden Unterschieden, denn natürlich streben Gymnasiasten häufiger das Abitur und ein Universitätsabschluss an als Schüler einer Sekundarschule, die häufiger ihre Schullaufbahn mit dem MSA beenden und eine Ausbildung anfangen wollen. Außerdem war bei den Sekundarschülern eine bessere Berufsvorbereitung festzustellen. Dieser Unterschied stellt jedoch keine alarmierende Erkenntnis dar, denn der Großteil der Schüler des Gymnasiums hat mindestens weitere zwei Jahre Zeit, um sich auf das Berufs- und Universitätsleben vorzubereiten. An der ISS sowie bei deren Schülern besteht eine viel höhere Dringlichkeit der Vorbereitung auf das Berufsleben, denn, wie mehrfach erwähnt, verlassen viele Schüler die ISS nach der 10. Klasse, um in das Berufsleben zu starten. Daher ist es zwangsläufig notwendig, spätestens gegen Ende der 10. Klasse zu wissen, welche Möglichkeiten einem offen stehen bzw. welche Tätigkeit man gerne ausüben würde.

Am Gymnasien besteht die Möglichkeit, berufsvorbereitende Veranstaltungen oder Karriere-Tage noch in der 11. Klasse und auch in höheren Jahrgängen anzubieten, so dass diesbezügliche Fragen für diesen Schultyp etwas verfrüht gewesen sein könnten.

Um klarere Aussagen zum Vergleich der Berufsvorbereitung an Gymnasien und ISS machen zu können, wären zusätzliche Befragungen Schüler höherer Jahrgänge nötig. Zudem könnte durch eine großflächigere Befragung von zufällig ausgewählten Schulen und Schülern eine externe Validität erreicht werden. Dadurch könnten Rückschlüsse auf alle Berliner Schulen und Schüler gezogen werden.

Ebenfalls sinnvoll wäre eine Wiederholung der durchgeführten Umfrage in regelmäßigen Abständen. Dadurch könnte untersucht werden, inwiefern die Berliner Schulstrukturreform zielführend ist, und ob Verbesserungen bei der Berufsvorbereitung der Schüler (einer der Hauptpunkte der Reform) erzielt werden konnten.

Bisher kann man konstatieren, dass zumindest die neu eingeführten Konzepte wie zum Beispiel das "Dualen Lernen" noch nicht sehr erfolgreich in den Schulalltag inte-

griert werden konnten. Denn an beiden untersuchten Schulen gab der Großteil der Befragten an, überhaupt nicht mit diesem Konzept vertraut zu sein.

# Anhang:

## A.1. Literatur- und Quellenverzeichnis

**AKREMI, L.; BAUR, N.; FROMM, S.:** Datenanalyse mit SPSS für Fortgeschrittene 1, Datenaufbereitung und uni- und bivariate Statistik, 3., überarbeitete und erweiterte Auflage. VS Verlag, 2011

**BACKHAUS, K.; ERICHSON, B.; PLINKE, W.; WEIBER, R.:** Multivariate Analysemethoden, Eine anwendungsorientierte Einführung, Dreizehnte, überarbeitete Auflage. Springer Verlag, 2011

**BLEYMÜLLER, J.:** Statistik für Wirtschaftswissenschaftler, 16. Auflage. Verlag Vahlen München, 2012

**BORTZ, J.:** Statistik für Human- und Sozialwissenschaftler, 6. Auflage. Springer Medizin Verlag, 2005

**CORTINA, K.; BAUMERT, J.; LESCHINSKY, A.; MAYER, K.; TROMMER, L.:** Das Bildungswesen in der Bundesrepublik Deutschland. Rowohlt Taschenbuch Verlag, 2008

**DZIUBAN, C.; SHIRKEY, E.:** When is a Correlation Matrix Appropriate for Factor Analysis? in: *Psychological Bulletin*, Vol. 81, Nr. 6, 1974

**FEND, H.:** Theorie der Schule, 1980

**FRIED, R.; KNIES, K.; KNÖLLMANN, C. - UNIVERSITÄT DORTMUND:** Skript "Multivariate Datenanalyse: <https://www.statistik.tu-dortmund.de/~fried/Multivariate/Skript/> (23.05.2014)

**HÄDER, M.:** Empirische Sozialforschung, Eine Einführung, 2., überarbeitete Auflage. VS Verlag, 2010

**HÄRDLE, W.; RÖNZ, B. - LADISLAUS VON BORTKIEWICZ LEHRSTUHL FÜR STATISTIK:** Fragebogenanalyse: [http://mars.wiwi.hu-berlin.de/mediawiki/mmstat\\_de/index.php/AQM-Einführung](http://mars.wiwi.hu-berlin.de/mediawiki/mmstat_de/index.php/AQM-Einführung) (18.05.2014)

**HINTZ, D.; PÖPPEL, K.; REKUS, J.:** Neues schulpädagogisches Wörterbuch. Juventa, 1993

**JACOB, R.; HEINZ, A.; DÉCIEUX, J.:** Umfrage, Einführung in die Methoden der Umfrageforschung, 3. Auflage. Oldenburg Verlag, 2013

**KAISER, H.; RICE, J.:** Little Jiffy, Mark IV, in: *Education and Psychological Measurement*, Vol. 34, 1974

**KIRCHHOFF, S.; KUHN, S.; LIPP, P.; SCHLAWIN, S.:** Der Fragebogen, Datenbasis, Konstruktion und Auswertung, 4, überarbeitete Auflage. VS Verlag, 2008

**KLOPP, E. - UNIVERSITÄT DES SAARLANDES:** Explorative Faktorenanalyse: <http://www.eric-klopp.de/texte/methoden/14-explorative-faktorenanalyse> (23.05.2014)

**RÜGER, B.:** Induktive Statistik, Einführung für Wirtschafts- und Sozialwissenschaftler, 3. Auflage. Oldenburg Verlage, 1996

**SCHNELL, R.; HILL, P.; ESSER, E.:** Methoden der empirischen Sozialforschung, 9. Auflage. Oldenburg Verlag, 2011

**SENATSWERWALTUNG FÜR BILDUNG, WISSENSCHAFT UND FORSCHUNG:** Die Berliner Schulstrukturreform: <http://www.berlin.de/imperia/md/content/sen-bildung/bildungspolitik/schulreform/schulstrukturpraesentation.pdf?start&ts=1260974103&file=schulstrukturpraesentation.pdf> (12.05.2014)

**SENATSWERWALTUNG FÜR BILDUNG, WISSENSCHAFT UND FORSCHUNG:** Bildungsfahrplan Berlin: <http://www.berlin.de/imperia/md/content/sen-bildung/bildungspolitik/schulreform/bildungsfahrplan.pdf?start&ts=1251469098&file=bildungsfahrplan.pdf> (12.05.2014)

**SENATSWERWALTUNG FÜR BILDUNG, WISSENSCHAFT UND FORSCHUNG:** Entwurf eines Leitbildes für das Gymnasium und die Integrierte Sekundarschule: [http://www.berlin.de/imperia/md/content/sen-bildung/bildungswege/entwurf\\_leitbild.pdf?start&ts=1353509468&file=entwurf\\_leitbild.pdf](http://www.berlin.de/imperia/md/content/sen-bildung/bildungswege/entwurf_leitbild.pdf?start&ts=1353509468&file=entwurf_leitbild.pdf) (12.05.2014)

**STEIN, M. - UNIVERSITÄT DUISBURG ESSEN:** Faktorenanalyse: <https://www.uni-due.de/imperia/md/content/soziologie/stein/faktorenanalyse.pdf> (23.05.2014)

**VON DETTEN, P.; FAUDE, O.; MEYER, T. - UNIVERSITÄT PADERBORN:** Leitfaden zur statistischen Auswertung von empirischen Studien: [http://dsg.uni-paderborn.de/fileadmin/dsg/ab3/Leitfaden\\_zur\\_statistischen\\_Auswertung\\_von\\_Studien.pdf](http://dsg.uni-paderborn.de/fileadmin/dsg/ab3/Leitfaden_zur_statistischen_Auswertung_von_Studien.pdf) (17.05.2014)

## A.2. Verzeichnis der Variablenabkürzungen

<b>age</b>	Alter	
<b>school_type_b</b>	0 = Sekundarschule	1 = Gymnasium
<b>sex_b</b>	0 = weiblich	1 = männlich
<b>avrgade</b>	Durchschnittsnote letztes Zeugnis	
<b>degree</b>	Angestrebter Schulabschluss	
<b>LKI, LKII, LKIII</b>	Wunsch-Leistungskurse	
<b>aftergrad</b>	Angestrebte Tätigkeit nach Schulabschluss	
<b>indI, indII, indIII</b>	Branchen, in denen sich der Befragte vorstellen könnte, zu arbeiten	
<b>oppind*</b>	Teiln. hat eine Vorstellung über berufl. Möglichkeiten in gewählten Branchen	
<b>thoughtgrad*</b>	Teiln. hat sich mit den Möglichkeiten nach dem Schulabschluss beschäftigt	
<b>job*</b>	Teiln. weiß, welchen Beruf er nach seinem Abschluss ausüben möchte	
<b>fear*</b>	Teiln. hat Angst vor dem Berufsleben	
<b>fearfun*</b>	Teiln. hat Angst einen Beruf zu haben, der keinen Spaß macht	
<b>prepschool*</b>	Teiln. wurde durch die Schule auf das Leben nach dem Abschluss vorbereitet	
<b>prepclass*</b>	Teiln. hat innerhalb des Unterrichts an Berufsvorbereitungsmaßn. teilgenommen	
<b>schoolusef*</b>	Teiln. fand die berufsbezogene Veranstaltungen in der Schule hilfreich	
<b>schoolinfo*</b>	Teiln. fand das Schulangebot ausreichend vorbereitend für den Berufseinstieg	
<b>prepprivat*</b>	Teiln. hat sich privat über Studien-, Ausbildungs- und Berufsmöglichkeiten informiert	
<b>prepmore*</b>	Teiln. denkt, er sollte sich mehr über seine Möglichkeiten informieren	
<b>ambitious*</b>	Teiln. schätzt sich als ehrgeizig ein	
<b>classbest*</b>	Teiln. zählt sich zu den besten in seiner Klasse	
<b>relatech*</b>	Teiln. denkt, er hat ein gutes Verhältnis zu seinen Lehrern	
<b>diligent*</b>	Teiln. schätzt sich als fleißig ein	
<b>organized*</b>	Teiln. schätzt sich als organisiert ein	
<b>goaloriented*</b>	Teiln. schätzt sich als zielorientiert ein	
<b>dedicated*</b>	Teiln. schätzt sich als engagiert ein	
<b>engfriends*</b>	Teiln. schätzt seine Freunde als engagiert und/oder zielorientiert ein	
<b>extracurr*</b>	Teiln. nimmt an Schulaktivitäten außerhalb des Unterrichts teil	
<b>dualschool*</b>	Teiln. kennt das Konzept Duale Schule	
<b>partdualsci*</b>	Teiln. nimmt an dem Konzept Duale Schule teil	

	Alle Fragen mit * sind Multiple-Choice-Fragen mit folgenden Antwortmöglichkeiten:			
	1 = trifft voll und ganz zu	2 = trifft eher zu	3 = trifft eher nicht zu	4 = trifft überhaupt nicht zu
<b>numberpriv</b>	Anzahl an privat durchgeführten Berufsvorbereitungsmaßnahmen			
<b>numberscl</b>	Anzahl an besuchten Berufsvorbereitungsmaßnahmen in Zsh. mit der Schule			
<b>intl, intl, intl</b>	Interessensgebiete des Befragten			
<b>preptotal</b>	Gesamteinschätzung zur Berufsvorbereitung des Befragten (Schulnotenskala)			
<b>mistakeincomplete_b</b>	0 = Fragebogen wurde korrekt ausgefüllt		1 = Fragebogen war unvollständig beantwortet oder fehlerhaft ausgefüllt	

## A.3. SPSS-Output

### a) Deskriptive Statistik:

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
age	227	13	18	15,69	,795
sex_b	220	0	1	,51	,501
school_type_b	230	0	1	,42	,495
degree_abi	230	0	1	,69	,465
degree_msa	230	0	1	,07	,262
degree_msamehr	230	0	1	,20	,401
degree_eBB	230	0	1	,03	,172
aftergrad_uni	230	0	1	,65	,477
aftergrad_ausbildung	230	0	1	,24	,430
aftergrad_berufseinstieg	230	0	1	,01	,093
aftergrad_sonstiges	230	0	1	,04	,194
preptotal	228	1	6	3,14	1,087
mistakeincomplete_b	230	0	1	,30	,459
Gültige Werte (Listenweise)	218				



### Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
age	130	15	18	16,02	,628
sex_b	126	0	1	,40	,491
degree_abi	133	0	1	,51	,502
degree_msa	133	0	1	,13	,335
degree_msamehr	133	0	1	,29	,457
degree_eBB	133	0	1	,05	,224
aftergrad_uni	133	0	1	,48	,502
aftergrad_ausbildung	133	0	1	,35	,480
aftergrad_berufseinstieg	133	0	1	,02	,122
aftergrad_sonstiges	133	0	1	,06	,239
preptotal	132	1	6	3,05	1,118
mistakeincomplete_b	133	0	1	,31	,464
Gültige Werte (Listenweise)	125				

a. school\_type\_b = 0

### Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
age	97	13	17	15,24	,774
sex_b	94	0	1	,66	,476
degree_abi	97	0	1	,93	,260
degree_msa	97	0	0	,00	,000
degree_msamehr	97	0	1	,07	,260
degree_eBB	97	0	0	,00	,000
aftergrad_uni	97	0	1	,89	,319
aftergrad_ausbildung	97	0	1	,09	,292
aftergrad_berufseinstieg	97	0	0	,00	,000
aftergrad_sonstiges	97	0	1	,01	,102
preptotal	96	1	6	3,28	1,033
mistakeincomplete_b	97	0	1	,29	,455
Gültige Werte (Listenweise)	93				

a. school\_type\_b = 1

### Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
age	108	14	17	15,72	,708
school_type_b	108	0	1	,30	,459
degree_abi	108	0	1	,56	,498
degree_msa	108	0	1	,07	,263
degree_msamehr	108	0	1	,31	,463
degree_eBB	108	0	1	,05	,211
aftergrad_uni	108	0	1	,56	,498
aftergrad_ausbildung	108	0	1	,27	,445
aftergrad_berufseinstieg	108	0	1	,01	,096
aftergrad_sonstiges	108	0	1	,06	,230
preptotal	107	1	6	3,21	1,206
mistakeincomplete_b	108	0	1	,30	,459
Gültige Werte (Listenweise)	107				

a. sex\_b = 0

### Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
age	112	13	18	15,66	,876
school_type_b	112	0	1	,55	,499
degree_abi	112	0	1	,81	,392
degree_msa	112	0	1	,07	,259
degree_msamehr	112	0	1	,11	,311
degree_eBB	112	0	1	,01	,094
aftergrad_uni	112	0	1	,74	,440
aftergrad_ausbildung	112	0	1	,21	,412
aftergrad_berufseinstieg	112	0	1	,01	,094
aftergrad_sonstiges	112	0	1	,02	,133
preptotal	111	1	6	3,05	,971
mistakeincomplete_b	112	0	1	,27	,445
Gültige Werte (Listenweise)	111				

a. sex\_b = 1

## b) Faktorenanalyse:

Korrelationsmatrix

	prepschool	prepclass	schoolusef	schoolinfo	ambitious	classbest	diligent	organized	goaloriented	engaged	oppind	thoughtgrad	job	preprivat
Korrelation	1,000	,500	,558	,411	-,013	-,053	,107	,019	-,005	,008	,038	,072	,139	,083
		1,000	,438	,422	-,068	-,047	,003	-,059	-,080	,011	,053	,073	,170	,104
			1,000	,493	,002	-,043	,056	-,061	,043	,039	-,045	,015	,003	-,024
				1,000	,022	,083	,030	-,049	-,027	,064	-,032	-,102	-,021	,030
					1,000	,391	,460	,353	,468	,555	,129	,188	,056	,244
						1,000	,286	,235	,250	,333	-,008	,026	-,119	,055
							1,000	,540	,404	,484	,126	,231	,198	,242
								1,000	,533	,410	,084	,196	,094	,196
									1,000	,561	,243	,387	,276	,270
										1,000	,250	,301	,179	,297
											1,000	,385	,422	,168
												1,000	,587	,426
													1,000	,277
														1,000

Anti-Image-Matrizen

		prepsc hool	prepcl ass	schoolu sef	schooli nfo	ambitio us	classb est	dilige nt	organiz ed	goalorie nted	dedica ted	oppin d	thought grad	job	prepri vat
Anti- Image- Kovarianz	prepschool	,577	-,165	-,217	-,076	-,002	,020	-,045	-,040	,028	,040	-,017	,001	-,044	-,032
	prepclass	-,165	,639	-,104	-,137	,029	-,017	,034	-,017	,075	-,030	-,010	-,001	-,083	-,058
	schoolusef	-,217	-,104	,551	-,184	,001	,065	-,038	,077	-,081	-,010	,050	-,050	,071	,083
	schoolinfo	-,076	-,137	-,184	,657	,000	-,086	,014	,017	,012	-,039	-,008	,093	-,011	-,048
	ambitious	-,002	,029	,001	,000	,577	-,137	-,115	,018	-,093	-,149	-,011	,005	,045	-,061
	classbest	,020	-,017	,065	-,086	-,137	,765	-,072	-,007	-,037	-,070	,027	-,027	,108	,050
	diligent	-,045	,034	-,038	,014	-,115	-,072	,562	-,221	,039	-,100	,019	,010	-,084	-,034
	organized	-,040	-,017	,077	,017	,018	-,007	-,221	,565	-,199	-,019	,031	-,004	,057	-,007
	goaloriented	,028	,075	-,081	,012	-,093	-,037	,039	-,199	,488	-,134	-,031	-,067	-,074	-,014
	engaged	,040	-,030	-,010	-,039	-,149	-,070	-,100	-,019	-,134	,510	-,084	-,024	,017	-,056
	oppind	-,017	-,010	,050	-,008	-,011	,027	,019	,031	-,031	-,084	,763	-,093	-,155	,032
	thoughtgrad	,001	-,001	-,050	,093	,005	-,027	,010	-,004	-,067	-,024	-,093	,519	-,233	-,185
	job	-,044	-,083	,071	-,011	,045	,108	-,084	,057	-,074	,017	-,155	-,233	,548	-,005
	preprivat	-,032	-,058	,083	-,048	-,061	,050	-,034	-,007	-,014	-,056	,032	-,185	-,005	,752
Anti- Image- Korrelation	prepschool	,744	-,272	-,385	-,123	-,003	,031	-,079	-,070	,053	,074	-,025	,001	-,078	-,049
	prepclass	-,272	,770	-,175	-,211	,048	-,024	,056	-,028	,135	-,052	-,015	-,002	-,141	-,084
	schoolusef	-,385	-,175	,668	-,305	,002	,100	-,068	,138	-,156	-,019	,077	-,093	,128	,128
	schoolinfo	-,123	-,211	-,305	,751	,000	-,122	,024	,028	,021	-,068	-,011	,160	-,019	-,068
	ambitious	-,003	,048	,002	,000	,845	-,206	-,201	,032	-,175	-,275	-,017	,008	,080	-,093
	classbest	,031	-,024	,100	-,122	-,206	,789	-,110	-,010	-,060	-,112	,035	-,044	,167	,066
	diligent	-,079	,056	-,068	,024	-,201	-,110	,802	-,392	,074	-,186	,029	,018	-,152	-,053
	organized	-,070	-,028	,138	,028	,032	-,010	-,392	,753	-,378	-,035	,048	-,007	,102	-,011
	goaloriented	,053	,135	-,156	,021	-,175	-,060	,074	-,378	,805	-,269	-,050	-,133	-,143	-,023
	dedicated	,074	-,052	-,019	-,068	-,275	-,112	-,186	-,035	-,269	,855	-,135	-,048	,032	-,091
	oppind	-,025	-,015	,077	-,011	-,017	,035	,029	,048	-,050	-,135	,822	-,147	-,240	,042
	thoughtgrad	,001	-,002	-,093	,160	,008	-,044	,018	-,007	-,133	-,048	-,147	,749	-,437	-,296
	job	-,078	-,141	,128	-,019	,080	,167	-,152	,102	-,143	,032	-,240	-,437	,683	-,008
	preprivat	-,049	-,084	,128	-,068	-,093	,066	-,053	-,011	-,023	-,091	,042	-,296	-,008	,812

a. Maß der Stichprobeneignung

RELIABILITY

```
/VARIABLES=oppind thoughtgrad job prepprivat
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA.
```

#### Zusammenfassung der Fallverarbeitung

	N	%
Gültig	222	96,5
Fälle    Ausgeschlossen	8	3,5
Gesamt	230	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen  
in der Prozedur.

#### Reliabilitätsstatistiken

Cronbachs Al- pha	Anzahl der Items
,703	4

RELIABILITY

```
/VARIABLES=prepschool prepclass schoolusef schoolinfo
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA.
```

#### Zusammenfassung der Fallverarbeitung

	N	%
Gültig	221	96,1
Fälle    Ausgeschlossen	9	3,9
Gesamt	230	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen  
in der Prozedur.

#### Reliabilitätsstatistiken

Cronbachs Al- pha	Anzahl der Items
,789	4

# RELIABILITY

```
/VARIABLES=ambitious classbest dilligent organized goaloriented dedicated
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA.
```

## Zusammenfassung der Fallverarbeitung

	N	%
Gültig	225	97,8
Fälle    Ausgeschlossen	5	2,2
Gesamt	230	100,0

a. Listenweise Löschung auf der Grundlage aller Variablen in der Prozedur.

## Reliabilitätsstatistiken

Cronbachs Alpha	Anzahl der Items
,809	6

### c. Clusteranalyse

**Zuordnungsübersicht**

Schritt	Zusammengeführte Cluster		Koeffizienten	Erstes Vorkommen des Clusters		Nächster Schritt
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	28	29	2,785	0	0	161
2	165	166	6,340	0	0	145
3	136	143	10,635	0	0	56
4	193	194	14,962	0	0	35
5	147	148	19,524	0	0	46
6	204	221	24,194	0	0	102
7	119	133	29,063	0	0	66
8	78	162	34,149	0	0	11
9	164	215	39,288	0	0	72
10	111	139	44,448	0	0	67
11	78	109	49,644	8	0	80
12	18	188	55,050	0	0	52
13	8	123	60,456	0	0	56
14	45	90	65,959	0	0	16
15	17	95	71,473	0	0	70
16	45	77	77,377	14	0	87
17	11	15	83,297	0	0	31
18	63	189	89,343	0	0	59
19	37	65	95,468	0	0	52
20	208	209	101,624	0	0	151
21	152	154	107,836	0	0	98
22	27	155	114,089	0	0	129
23	97	108	120,370	0	0	112
24	21	73	126,780	0	0	33
25	34	35	133,280	0	0	137
26	96	212	139,815	0	0	114
27	67	88	146,433	0	0	114
28	56	177	153,089	0	0	134
29	86	192	159,832	0	0	142
30	40	134	166,586	0	0	134
31	11	13	173,365	17	0	79
32	38	116	180,170	0	0	125
33	21	64	187,066	24	0	128
34	121	140	193,986	0	0	106

35	137	193	200,965	0	4	165
36	85	196	207,969	0	0	121
37	110	180	214,972	0	0	81
38	106	219	222,003	0	0	102
39	31	107	229,035	0	0	55
40	61	87	236,160	0	0	95
41	19	43	243,455	0	0	149
42	183	230	250,823	0	0	68
43	20	44	258,229	0	0	94
44	75	179	265,649	0	0	85
45	47	197	273,153	0	0	106
46	147	190	280,775	5	0	124
47	3	138	288,404	0	0	122
48	124	142	296,060	0	0	109
49	141	210	303,720	0	0	97
50	94	195	311,390	0	0	124
51	9	30	319,078	0	0	138
52	18	37	326,767	12	19	91
53	105	127	334,517	0	0	132
54	53	54	342,312	0	0	132
55	25	31	350,133	0	39	138
56	8	136	358,004	13	3	150
57	169	229	365,883	0	0	148
58	128	181	373,764	0	0	88
59	62	63	381,774	0	18	73
60	156	171	389,806	0	0	135
61	2	24	397,840	0	0	100
62	163	173	405,979	0	0	133
63	178	224	414,135	0	0	135
64	51	81	422,292	0	0	116
65	10	48	430,528	0	0	137
66	119	202	438,799	7	0	141
67	111	176	447,194	10	0	107
68	117	183	455,650	0	42	107
69	1	12	464,206	0	0	113
70	17	58	472,766	15	0	91
71	89	199	481,645	0	0	84
72	164	216	490,524	9	0	144
73	59	62	499,514	0	59	154
74	7	57	508,611	0	0	116



75	198	220	517,892	0	0	144
76	4	186	527,188	0	0	126
77	42	76	536,514	0	0	113
78	149	170	545,851	0	0	172
79	11	122	555,197	31	0	180
80	78	103	564,619	11	0	130
81	110	113	574,046	37	0	151
82	100	158	583,519	0	0	136
83	50	146	593,019	0	0	122
84	68	89	602,529	0	71	154
85	75	145	612,086	44	0	150
86	118	161	621,686	0	0	118
87	41	45	631,298	0	16	139
88	99	128	640,925	0	58	131
89	80	207	650,637	0	0	121
90	72	104	660,387	0	0	111
91	17	18	670,491	70	52	139
92	114	132	680,616	0	0	123
93	46	74	690,772	0	0	140
94	20	84	700,950	43	0	128
95	55	61	711,288	0	40	130
96	203	227	721,666	0	0	127
97	141	157	732,191	49	0	164
98	152	223	742,887	21	0	171
99	150	174	753,807	0	0	147
100	2	26	764,911	61	0	176
101	69	70	776,040	0	0	167
102	106	204	787,193	38	6	141
103	167	218	798,349	0	0	133
104	98	131	809,505	0	0	160
105	49	185	820,759	0	0	131
106	47	121	832,033	45	34	127
107	111	117	843,437	67	68	142
108	66	91	854,909	0	0	152
109	115	124	866,419	0	48	143
110	71	214	878,048	0	0	174
111	5	72	889,761	0	90	129
112	97	206	901,512	23	0	147
113	1	42	913,356	69	77	152
114	67	96	925,613	27	26	169

115	23	32	937,981	0	0	178
116	7	51	950,378	74	64	168
117	126	159	962,785	0	0	162
118	118	120	975,402	86	0	188
119	16	187	988,433	0	0	126
120	33	83	1001,523	0	0	155
121	80	85	1014,617	89	36	158
122	3	50	1027,723	47	83	159
123	92	114	1040,931	0	92	166
124	94	147	1054,209	50	46	163
125	38	211	1067,626	32	0	177
126	4	16	1081,310	76	119	170
127	47	203	1095,068	106	96	143
128	20	21	1108,985	94	33	173
129	5	27	1123,079	111	22	140
130	55	78	1137,197	95	80	153
131	49	99	1151,516	105	88	148
132	53	105	1165,956	54	53	163
133	163	167	1180,409	62	103	136
134	40	56	1195,128	30	28	153
135	156	178	1209,909	60	63	157
136	100	163	1224,939	82	133	187
137	10	34	1240,058	65	25	164
138	9	25	1255,379	51	55	159
139	17	41	1270,713	91	87	170
140	5	46	1286,283	129	93	161
141	106	119	1302,155	102	66	171
142	86	111	1318,124	29	107	167
143	47	115	1334,251	127	109	165
144	164	198	1350,543	72	75	179
145	135	165	1366,867	0	2	162
146	182	228	1383,204	0	0	156
147	97	150	1399,601	112	99	175
148	49	169	1416,973	131	57	175
149	14	19	1434,410	0	41	158
150	8	75	1452,110	56	85	177
151	110	208	1469,887	81	20	166
152	1	66	1487,719	113	108	172
153	40	55	1505,556	134	130	182
154	59	68	1523,531	73	84	169

155	33	213	1541,598	120	0	185
156	182	201	1559,724	146	0	189
157	156	175	1578,203	135	0	183
158	14	80	1596,694	149	121	186
159	3	9	1615,216	122	138	168
160	98	112	1634,435	104	0	178
161	5	28	1653,771	140	1	188
162	126	135	1673,349	117	145	183
163	53	94	1693,940	132	124	179
164	10	141	1714,686	137	97	184
165	47	137	1735,666	143	35	181
166	92	110	1757,087	123	151	192
167	69	86	1779,018	101	142	182
168	3	7	1801,433	159	116	186
169	59	67	1823,882	154	114	187
170	4	17	1847,186	126	139	173
171	106	152	1870,970	141	98	190
172	1	149	1894,779	152	78	176
173	4	20	1919,151	170	128	184
174	71	225	1943,611	110	0	189
175	49	97	1968,197	148	147	191
176	1	2	1994,408	172	100	180
177	8	38	2020,741	150	125	181
178	23	98	2049,996	115	160	185
179	53	164	2079,496	163	144	190
180	1	11	2110,381	176	79	198
181	8	47	2141,457	177	165	191
182	40	69	2174,044	153	167	193
183	126	156	2208,297	162	157	194
184	4	10	2242,907	173	164	195
185	23	33	2277,932	178	155	200
186	3	14	2313,251	168	158	197
187	59	100	2348,598	169	136	196
188	5	118	2384,874	161	118	195
189	71	182	2422,551	174	156	200
190	53	106	2463,983	179	171	192
191	8	49	2510,470	181	175	193
192	53	92	2557,155	190	166	194
193	8	40	2624,214	191	182	196
194	53	126	2695,196	192	183	199

195	4	5	2766,505	184	188	197
196	8	59	2840,260	193	187	199
197	3	4	2918,604	186	195	198
198	1	3	3003,719	180	197	201
199	8	53	3109,275	196	194	201
200	23	71	3242,345	185	189	202
201	1	8	3488,813	198	199	202
202	1	23	3806,333	201	200	0

## A.4. Muster-Fragebogen

### Umfrage zur Vorbereitung auf das Berufsleben während der Schulzeit

Alter: \_\_\_\_\_

Geschlecht: ☐ w ☐ m

Notendurchschnitt letztes Zeugnis:

☐ 1,0 bis unter 2,0 ☐ 2,0 bis unter 3,0 ☐ 3,0 bis unter 4,0 ☐ schlechter als 4,0

1. Ich besuche ein(e):		
<input type="checkbox"/> Gymnasium	<input type="checkbox"/> Sekundarschule	
2. Ich strebe folgenden Abschluss an:		
<input type="checkbox"/> Abitur <input type="checkbox"/> MSA <input type="checkbox"/> MSA, aber ich würde gerne später Abi oder Fachabi machen <input type="checkbox"/> eBB (früher: erweiterter Hauptschulabschluss)		
3. Diese Leistungskurse belege ich/möchte ich belegen:		
I.	II.	III. (falls zutrifft)
4. Ich strebe eine(n) .... an.		
<input type="checkbox"/> Universitätsabschluss <input type="checkbox"/> Ausbildung <input type="checkbox"/> direkten Berufseinstieg <input type="checkbox"/> Sonstiges, nämlich _____		
5. Ich könnte mir vorstellen, in diesen Branchen zu arbeiten (max 3 wählen):		
<input type="checkbox"/> Automobilindustrie <input type="checkbox"/> Chemieindustrie <input type="checkbox"/> Banken und Versicherungen <input type="checkbox"/> Medien (TV, Radio, Zeitung, usw.) <input type="checkbox"/> Lebensmittelindustrie <input type="checkbox"/> Informations-/Kommunikationstechnologie <input type="checkbox"/> Elektronikindustrie <input type="checkbox"/> Maschinen- und Anlagebau <input type="checkbox"/> Energiebranche <input type="checkbox"/> Logistik <input type="checkbox"/> Tourismus- und Hotelbranche <input type="checkbox"/> Gesundheitsbranche <input type="checkbox"/> Bildung <input type="checkbox"/> Pflegebereich <input type="checkbox"/> Eventmanagement		

		trifft voll und ganz zu	trifft eher zu	trifft eher nicht zu	trifft überhaupt nicht zu
6.	Ich habe eine ungefähre Vorstellung, welche beruflichen Möglichkeiten ich in den gewählten Branchen habe:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7.	Ich habe mir bereits Gedanken gemacht, was ich nach meinem Schulabschluss machen möchte:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8.	Ich weiß bereits, welchen Beruf ich ausüben möchte:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9.	Ich habe Angst vor dem Berufsleben:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	Ich habe Angst, einen Beruf zu haben, der mir keinen Spaß macht:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11.	Ich wurde durch die Schule auf das Leben nach meinem Schulabschluss vorbereitet (durch bspw. Gastvorträge, Messe- und Betriebsbesuche, usw.):	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12.	Auch innerhalb des normalen Lehrbetriebs finden/fanden Berufsvorbereitungsmaßnahmen statt (z.B. in Arbeitslehre oder anderen Fächern):	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

13.	Ich finde, dass alle berufsbezogenen Veranstaltungen, die ich in der Schule besuche, für meine zukünftige Berufsauswahl hilfreich sind:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14.	Das Informations-/Veranstaltungsangebot der Schule genügt, um für den Uni-/Berufseinstieg vorbereitet zu sein:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

15.	Ich habe mich auch privat über mögliche Studiengänge bzw. Ausbildungs- und Berufsmöglichkeiten informiert:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16.	Eigentlich sollte ich mich mehr über die Möglichkeiten, die ich nach meinem Schulabschluss habe, informieren:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17.	Ich bin ehrgeizig:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18.	Ich gehöre zu den Besten in meiner Klasse:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19.	Prinzipiell habe ich ein gutes Verhältnis zu meinen Lehrern:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20.	Ich bin fleißig:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21.	Ich bin organisiert:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22.	Ich bin zielorientiert:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23.	Ich bin engagiert:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24.	Ich habe Freunde, die engagiert und/oder zielorientiert sind:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25.	Ich nehme an Schulaktivitäten außerhalb des Unterrichts teil (AG's, Clubs, Chor, Jugend forscht, usw.):	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26.	Ich bin mit dem Konzept "Duale Schule" vertraut:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27.	Ich war/bin selbst Teilnehmer an der "Dualen Schule":	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

28. Anzahl an Veranstaltungen zur Berufsvorbereitung, die privat von mir besucht wurden:			
<input type="checkbox"/> keine	<input type="checkbox"/> 1-3	<input type="checkbox"/> 4-6	<input type="checkbox"/> mehr als 6
29. Anzahl an besuchten Veranstaltungen zur Berufsvorbereitung außerhalb des Unterrichtsbetriebs und von der Schule organisiert (Wandertage zu Unternehmen, Berufsinformationszentrum, usw.):			
<input type="checkbox"/> keine	<input type="checkbox"/> 1-3	<input type="checkbox"/> 4-6	<input type="checkbox"/> mehr als 6
30. Aus der folgenden Liste interessieren mich am meisten (max. 3 auswählen):			
<input type="checkbox"/> Sport	<input type="checkbox"/> Musik & Instrumente	<input type="checkbox"/> Freunde	
<input type="checkbox"/> Film & Fernsehen	<input type="checkbox"/> Videospiele	<input type="checkbox"/> Kunst & Kultur	
<input type="checkbox"/> Reisen	<input type="checkbox"/> Shoppen	<input type="checkbox"/> Bildung	
<input type="checkbox"/> Personen <input type="checkbox"/> Politik	<input type="checkbox"/> Wirtschaft	<input type="checkbox"/> Nachrichten	
<input type="checkbox"/> Stars/berühmte Personen	<input type="checkbox"/> Sonstiges, nämlich _____	<input type="checkbox"/> Karriere	
31. So gut fühle ich mich insgesamt auf das Uni- und Berufsleben vorbereitet:			
<input type="checkbox"/> 1 (sehr gut)	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
<input type="checkbox"/> 5	<input type="checkbox"/> 6 (überhaupt nicht)		

Alle Angaben werden anonym und vertraulich behandelt. Falls eine Person durch die Angaben im Fragebogen identifiziert werden könnte, werden die Angaben im Sinne des Datenschutzes behandelt und falls nötig, werden bestimmte Beobachtungen aus der Analyse herausgenommen.

Hiermit erkläre ich, Nico Keskin, dass die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel von mir angefertigt wurde.

Die Prüfungsordnung ist mir bekannt. Ich habe in meinem Studienfach bisher keine Bachelorarbeit eingereicht bzw. diese nicht endgültig bestanden.

Nico Keskin

Berlin, 8. Juli 2014